

VISIT...

LANZAROTE
Caliente.COM

COMMUNICATION NETWORKS

VOL. II

THE CLASSICAL THEORY OF LONG LINES,
FILTERS AND RELATED NETWORKS

BY

ERNST A. GUILLEMIN, PH.D.

*Assistant Professor of Electrical Engineering,
Massachusetts Institute of Technology*

NEW YORK

JOHN WILEY & SONS, INC.

LONDON: CHAPMAN & HALL, LIMITED

1935

ETHERFORCE

COPYRIGHT, 1935,
BY ERNST A. GUILLEMIN

All Rights Reserved

*This book or any part thereof must not
be reproduced in any form without the
written permission of the publisher.*

Printed in U. S. A.

THE HADDON CRAFTSMEN, INC.
CAMDEN, N. J.

PREFACE

Specifically, the present volume is intended to present a thorough treatment of the transmission line as a communication facility, leading quite naturally into the field of filter theory and its related problems. More generally, however, the discussions pertain to the field of network theory as a whole, and apply to the power as well as to the communications aspect. The communications viewpoint is chosen for the more detailed discussions because it affords at present a more suitable vehicle for illustrating the fundamental principles. This is due in part to the fact that the consideration of network functions versus frequency, which is so essential to the study of network properties, is more vital to the communications aspect. Then, too, the problem of synthesis, which has been an important motivating influence in the enlargement of our views and processes with regard to network theory, not only had its inception in the field of communications but owes its development almost wholly to workers in that field. Nevertheless, these ideas and principles are too general in nature to remain confined to one field of application, and are more recently beginning to make their appearance in the treatment also of power problems.

The content of the chapters dealing with the properties of driving-point and transfer impedances and admittances, the fundamental forms of two- and four-terminal networks in the reactive or dissipative cases, the methods of synthesis of networks from prescribed impedance or admittance functions, the subjects of network equivalence, reciprocity, duality, and the use of linear transformations should become as helpful to the power engineer as they have been indispensable to the development of the present communications aspect of network behavior and design.

The synthesis point of view in network design has been with us for hardly more than a decade. During this brief period, particularly during the latter half, the activity has been very marked, and the contributions of ideas and viewpoints have been numerous and diverse. Some of these are more fundamental and permanently useful than others. Although, at present, it is difficult to judge the ultimate utility of much of this material, an attempt has been made to select for presentation only such contributions as are thought to be of more permanent value, and to weave these into the thread of the discussions in

such a way as to obtain, as nearly as possible, a logically connected development of thought.

In the preparation of the material for this volume I have been materially aided by Dr. O. Brune, with whom I have been associated in the teaching of the subject matter on network synthesis. His many suggestions and valuable council have largely influenced the form of treatment in those parts of the manuscript dealing with the synthesis point of view. I am grateful also to Mr. H. W. Bode, of the Bell Telephone Laboratories, for his kindness in making available to me the material dealing with his contributions to the subject of filter design as well as numerous other points of view on the subject at large. With reference to the relation of the engineering formulation of the line problem to the more rigorous electromagnetic field aspects I am indebted to Dr. J. A. Stratton for helpful criticism and suggestions. To Prof. E. L. Bowles and Prof. D. C. Jackson I wish to express my appreciation for their encouragement and cooperative attitude.

E. A. G.

CAMBRIDGE, MASS.

September, 1935

CONTENTS

	PAGE
CHAPTER I	
THE ENGINEERING FORMULATION OF THE LONG LINE PROBLEM.....	1
Distinction between distributed and lumped constant systems. Historical background regarding the theoretical treatment of the problem. The concept of guided waves. The semi-rigorous treatment of the long line problem. The engineering treatment of the long line problem. Discussion of the line parameters. Bibliography.	
CHAPTER II	
THE STEADY-STATE SOLUTION TO THE LONG LINE EQUATIONS.....	33
The general linear character of the equations. Formal solution by separation of variables. Direct solution for the steady state. Boundary conditions and the evaluation of integration constants. Wave character of the solution. Characteristic impedance. Wavelength and phase velocity. Hyperbolic form. Symmetrical forms. Neglected dissipation. Standing waves. Polar plots. Odd and even quarter-wavelength lines. Effect of attenuation. General circuit parameters. Input impedance. Voltage, current and power ratios. Attenuation units.	
CHAPTER III	
PROPAGATION AND CHARACTERISTIC IMPEDANCE FUNCTIONS OF THE LONG LINE.....	80
Ideal behavior. Separation of the propagation function. Discussion of the attenuation and phase functions. Amplitude and phase distortion. Tolerances and frequency limits. Phase and group velocity. Discussion of the characteristic impedance function. Net effect of various tolerances. Approximations in the consideration of the attenuation and phase functions. Approximations in the consideration of the characteristic impedance function.	
CHAPTER IV	
CHARACTERISTICS OF FOUR-TERMINAL NETWORKS.....	132
Significance of the problem. Possible relationships between voltages and currents. Derivation of fundamental relations. Symmetrical networks. The use of matrix algebra in the study of four-terminal network behavior. Ideal transformers and transformers without loss. Discussion of several basic structures. Characteristic impedance and propagation functions. Transmission and attenuation properties of networks. Propagation and characteristic impedance functions of several common structures.	

CHAPTER V

THEOREMS REGARDING DRIVING-POINT IMPEDANCES AND AN EXTENSION TO TWO-TERMINAL PAIRS IN THE REACTIVE CASE.....	184
Foster's reactance theorem. Function theoretical point of view. Special cases. Illustrations. Least number of elements. Cauer's extension of Foster's theorem. Equivalence and reciprocity. Bartlett's reciprocation theorem. Extension of the reactance theorem to dissipative cases. The reactance theorem extended to a two-terminal pair.	

CHAPTER VI

ENERGY FUNCTIONS AND LINEAR NETWORK TRANSFORMATIONS.....	222
General remarks. Lagrange's equations for the dynamic equilibrium of linear systems. The driving-point impedance function and its properties in terms of the energy functions. Transformations which keep an impedance invariant. Transformation by matrix manipulation. Extension to more pairs of terminals. Illustrations. Transformation to normal coordinates. The principle of duality. Equivalent networks by means of their dual representation. Inverse networks.	

CHAPTER VII

ARTIFICIAL AND LUMP-LOADED LINES.....	256
Lumped-section equivalents to the long line. The use of lattice structures. Lump-loaded lines. Cut-off frequency of lump-loaded lines.	

CHAPTER VIII

THE NON-DISSIPATIVE UNIFORM LADDER STRUCTURE.....	279
Formulation of the problem. The mid-series terminated structure. Physical interpretation. The behavior in the transmission range. The behavior in the attenuation range. Adjustment for best terminal conditions.	

CHAPTER IX

CONVENTIONAL FILTER THEORY.....	299
Basic design considerations. The constant- k type. Design procedure for the constant- k type. The use of half-sections. Several common classes of constant- k filters. Shortcomings of the constant- k type. The m -derived type. The composite filter. Repeated derivations. Reflection effects in the composite filter. Fractional terminations. Impedance correction. Other conventional filter types.	

CHAPTER X

FILTER DESIGN METHODS BASED UPON THE LATTICE STRUCTURE.....	378
Introductory remarks. Conditions for transmission and attenuation. The approximation problem. Some useful lattice properties. Impedance and index functions for various filter classes. Frequency transformations. Cauer's method of parameter determination. Illustrations. Bode's method of parameter determination. Ladder developments and other lattice equivalents. Dissymmetrical filters. Reflection losses. The effect of incidental dissipation. Concluding remarks.	

CHAPTER XI

THE TRANSIENT BEHAVIOR OF FILTERS.....	461
The Fourier integral. Relation to contour integrals. Response of linear systems. The response of idealized selective networks. Phase distortion. Simultaneous amplitude and phase distortion. The contour integral method of treatment. The response functions of physical networks. Concluding remarks.	

CHAPTER XII

SIMULATIVE AND CORRECTIVE NETWORKS.....	508
General remarks. Amplitude and phase distortion corrective networks. Compensating and simulating networks.	

CHAPTER XIII

THE TRANSIENT BEHAVIOR OF LONG LINES.....	554
Introductory remarks. The solution in terms of normal functions. The expansion problem. Terminal conditions leading to orthogonality. Terminal conditions to which Rayleigh's method applies. Solution of the determinantal equation. Evaluation of integrals. Applications. The properly terminated line.	

APPENDIX. LIST OF SYMBOLS.....	578
--------------------------------	-----

INDEX.....	581
------------	-----

COMMUNICATION NETWORKS

VOL. II

CHAPTER I

THE ENGINEERING FORMULATION OF THE LONG LINE PROBLEM

1. **Distinction between distributed and lumped constant systems.** The transmission of intelligence by electrical means is generally accomplished in either of two ways. One usually employs a pair of parallel wires; the other relies more upon the phenomenon of wave propagation through free space or along the earth's surface. Although the second is physically more involved, and could hardly be classified under the head of a circuit problem, the two methods have much in common from the standpoint of the fundamental principles employed in their analysis. The reason for this lies in the fact that both means involve the transmission of electrical energy through continuous media. The parallel conductor problem, which will receive a large portion of our attention in this volume, is, from the engineering standpoint, closely associated with the principles of ordinary circuit theory already discussed. It is usually classified as a circuit problem of a higher order of complexity, and its treatment, therefore, is logically assumed to follow that of the lumped-constant network rather than precede it. Although this order of treatment may be justified on the basis of the mathematical aspect, it is rather illogical with regard to the fundamental physical principles involved. These have to do with that which distinguishes the networks considered heretofore from the type of network referred to as the transmission line. The latter is called a *distributed-constant* system; the former is designated as one involving *lumped constants*. Our object in the present section is to show that the distributed system is really the common ground or starting point, and that the so-called lumped-constant network is an approximation which the engineering treatment of network theory finds justified and useful. It is essential to obtain as clear a picture as possible of the basis upon which the distinction between lumped and distributed systems is made before taking up a more detailed consideration of the latter.

In the treatment of ordinary lumped networks we made an assumption which, at that time, was not specifically pointed out to the reader. For example, when we considered a battery suddenly impressed upon a coil, we assumed that this impressed force was felt simultaneously throughout the coil, and that the current which flowed in response to this force was at every instant identical in all parts of the coil. Although this assumption seems justified in view of the small physical dimensions involved, it is nevertheless not an accurate description of the phenomenon. The impressed force is not simultaneously felt throughout the coil, nor is the resulting current the same in all parts of the coil at any instant. The impressed force is propagated with a finite velocity so that various parts of the coil experience its effect at slightly different times, depending upon the rate of propagation and the physical dimensions involved. The current which flows is, therefore, also not quite the same at all points for any given instant. In general, we expect that if the coil or any combination of coils, condensers, and resistances, for that matter, are very large in size, the assumption that the response, as well as the disturbing force, is simultaneous throughout, becomes untenable. In addition to the physical size of the system, its physical nature, as well as that of the driving force, also influence the magnitude of the error involved in such an assumption. If the latter is justified, the solution of the problem is much simpler, as the reader can easily appreciate. We then say that we are dealing with a lumped system. When conditions are such that the assumption introduces too large an error, then the distributed character must be partly or wholly taken into account. Ordinarily, in the treatment of the transmission line the distributed character is only partly taken into account. The nature of this situation and the justification for the usual procedure will now be discussed.

For the engineer the above distinction between lumped and distributed systems is entirely too vague. It is necessary, therefore, that we formulate a more concrete definition, or at least a basis for such a definition. Although such a formulation seems apropos at the present, we shall have to ask the indulgence of the reader in tolerating the introduction of certain terms, which he cannot be expected to appreciate fully at this time, and which certainly will not be treated in detail until very much later in this discussion. We assume, however, that the reader has been placed in this position before, and has become somewhat accustomed to it. Unfortunately, the presentation of this subject necessitates this sort of procedure at various times.

We shall see later that any disturbance may be considered as composed of a linear superposition of simple harmonic components, which

may also be called "frequency components" since each simple harmonic function has a definite frequency whereby it may be characterized. If the disturbance is in the nature of a traveling wave of arbitrary shape, then these components are referred to as component waves. We shall also see that each component is propagated with a definite velocity, usually in the vicinity of the velocity of light. Each component wave is not only a simple harmonic function of distance but, at a fixed point, presents a simple harmonic function of time as well. This will be clear if the reader will visualize a water wave having a simple harmonic shape and traveling at a constant velocity. A cork floating on the wave and restrained from moving horizontally will be observed to bob up and down with simple harmonic motion. It should also be clear that the velocity of propagation of the harmonic wave, the length of the wave (i.e., the distance from crest to crest or trough to trough), and the frequency with which the cork bobs up and down, are not independent quantities but bear a definite relationship to one another. The cork bobs up and down once for each passage of one wave length. If we call the temporal period of the cork τ , the wave length λ , and the velocity v , then we see that:

$$\frac{\lambda}{\tau} = v = \lambda \cdot f \quad (1)$$

where f is the frequency with which the cork bobs.

Thus we find that to each "component frequency," in a given disturbance, there corresponds a definite wave length. Throughout this length the component wave changes in phase by 2π radians. The criterion as to whether or not a physical system may be considered lumped at a point may now be formulated somewhat more definitely as follows: If the shortest essential wave length in the disturbance (as determined from the highest essential frequency by the relation (1)) is very large compared with the maximum physical dimension of the system acted upon (say 50 or 100 times as large), then the maximum phase difference in this component wave at any two points in the system must be very small (not in excess of $1/50$ or $1/100$ of 2π radians). This is identical with saying that the component wave has the same instantaneous value at all points in the system to within a maximum error dependent upon this small phase difference. If this error is small enough to be negligible under the circumstances, then the system may be considered lumped at a point.

The rigorous application of this criterion may be rather difficult. It necessitates a knowledge of the "largest physical dimension" (which need not be the straight line from end to end of the apparatus) and also

ETHERFORCE

the velocity of propagation of the shortest component wave along this path, both of which can only be estimated. However, we wish here to impress upon the reader the significance of the criterion rather than the manner of its application.

As a numerical illustration, let us consider a system which is to operate in connection with audio frequencies up to 5,000 cycles per second. Assuming that the velocity of propagation equals that of light, which is $3 \cdot 10^{10}$ cm per sec, this corresponds to a shortest wave length of $6 \cdot 10^6$ cm or 60 kilometers, which is about 37 miles. If the system consists of an arrangement of coils and condensers in a laboratory set-up or the like, it would certainly be safe to say that we may consider it lumped even though our above estimate is off by a factor of 10. On the other hand, the system may be a telephone line 50 or 100 miles long. Then we certainly could not consider the problem under the head of a lumped-constant affair. In cases which are less clear-cut than this, or where there may be more or less doubt, the criterion must be applied with more rigor, or the question may be referred to the laboratory. We might mention in this connection that in using frequencies such as occur in short-wave radio work, involving wave lengths of 10 meters or less, we are beginning to reach the limit within which our apparatus may be considered lumped. The manner in which such circuits may be effectively treated analytically still remains to be determined.

It may be well at this time to point out to the reader that where the lumped-constant treatment is no longer valid, it becomes equally difficult to adhere to our usual concepts of inductance and capacitance. For example, the inductance of a closed loop of wire is defined as the quotient of the flux linking the loop and the current flowing in the wire, assuming the wire diameter to be negligible. If the current is constant, this is a perfectly definite quantity. If the loop is excited by a very high-frequency source, however, then the current at any instant may vary considerably throughout the loop. Hence, although we could measure the total flux linking the loop at that instant, we would be at a loss as to what value of current to divide by. We could no longer assign an inductance to the loop. A similar situation applies to capacitance, which is usually defined as the ratio of the total charge carried by a pair of conductors to the potential difference between them. If the frequency of the source of excitation is very high so that the wave length is comparable with the dimensions of the conductors, then at a given instant the potential difference between them will vary considerably from point to point. In fact, we can no longer speak of a voltage between any two fixed points, because the line integral of

electric field intensity by which this potential difference is defined is no longer independent of the path of integration.

Thus we recognize that the so-called lumped-constant treatment of networks is always an approximation except in the case of steady-state behavior with constant applied forces when the laws of electro- and magneto-statics rigidly apply. Whenever the exciting forces are variable with time, as they are in almost every network problem, the electric and magnetic fields are no longer stationary, and the concepts of inductance and capacitance, which are so useful to the engineer, are no longer tenable in the light of rigid fundamental electromagnetic theory. Fortunately, however, for those frequencies with which we have to deal most commonly, the error made by disregarding this fact is negligible for practical purposes. We call such fields *quasi-stationary* to distinguish them from the very rapidly changing fields for which the error is too large to be neglected. Naturally enough, there is no sharp line of demarcation between those fields which may be considered quasi-stationary and those which can no longer be so treated. The one continuously merges with the other, and it is strictly a question of allowable tolerances as to whether the one or the other point of view may be taken.

In communication work, two general classes of problems fall in the quasi-stationary category. One is that involving the so-called lumped-constant networks, which were treated in the first volume. The other is the long transmission line problem which will occupy a large part of our attention here. The long-line problem as it presents itself in communication work may still be treated as a quasi-stationary field problem, but not as a lumped-constant problem, for reasons already pointed out. The concepts of inductance and capacitance may still be used as approximations although in a modified sense which we shall describe in more detail below. Roughly we designate such problems as the long line as distributed-constant problems to distinguish them from those where the parameters may be considered lumped. We speak of uniformly distributed inductance, capacitance, resistance, and leakage conductance as the parameters which characterize the uniform long line. The form of solution based upon such parameters is often referred to as the *engineering solution*. It carries with it a number of tacit assumptions and approximations, as may be expected from the circumstance that a rather complex system such as the long line or cable is completely described by specifying just four quantities. The justification for its introduction and use lies in the tremendous simplification which arises from this form of representation. It is safe to say that the long line as a communication channel would still be in

ETHERFORCE

the experimental stage if it had not been for the development of the engineering form of solution, so great are the benefits derived from the resulting simplification. At the same time it would be a poor policy for the engineer to ignore entirely the more rigorous aspect of this very important problem as well as the process of evolution out of which the engineering formulation was born. An appreciation of this situation is useful not only for the sake of its general informational value but also because it affords a criterion whereby we may approximately determine the range of validity of the results derivable from such a method of solution. It is just as important to know the range within which a solution may be expected to be at least approximately correct as it is to determine such a solution. In fact, the solution would be quite useless or even misleading unless we could determine the corresponding limits within which it might safely be applied. Most of the following discussion is given for the purpose of elaborating upon this point.

2. Historical background regarding the theoretical treatment of the problem. Long lines were used in electrical communications for many years before any clear conception was attained regarding their theoretical behavior. The electric telegraph began to develop commercially in the early 1840's. The first attempt at an analysis of the propagation of electrical energy through a long uniform circuit was made by Sir W. Thomson (Lord Kelvin) in 1855.¹ His analysis, however, was limited in application to the single wire conductor with concentric return such as was used in undersea cable circuits. The return may actually have been through the sea water itself, but electrically this would amount to practically the same thing. The land circuits in those days were all of the single conductor with earth-return type, to which Thomson's analysis did not apply because he did not consider the inductive effect, which in such a line very materially affects the resulting behavior. His treatment takes into account only the capacitance and resistance, which, as we shall see later, is sufficient only in cable circuits and for these only over a limited frequency range.

To G. Kirchhoff must be given the credit for first introducing the inductive effect into the analysis of the long-line behavior. The results of his investigation are given in two papers published in 1857.² It appeared later that W. Weber³ had independently solved the same

¹ "On the Theory of the Electric Telegraph," *Math. Phys. Papers* 2, p. 61, Proc. Royal Society, London, 1855.

² "Über die Bewegung der Elektrizität in Drähten," *Pogg. Ann.* 100, p. 193.

"Über die Bewegung der Elektrizität in Leitern," *Pogg. Ann.* 102, p. 529.

³ *Pogg. Ann.* 100, p. 351, 1857.

problem but had withheld publication until he could include the results of certain experimental verifications which he was undertaking in collaboration with R. Kohlrausch. This duplication of effort shows that the problem must have attracted considerable attention at that time. It is also interesting to note that the results of these investigations agree in every essential detail. The premises, however, are unhappily of such an idealized nature as to make the results of little practical value. This is due to the fact that both investigators assume the single wire to be so far removed from other conductors that it may be considered isolated. For this reason the inductive effect does not appear in its true light so far as practical circuits are concerned. The single conductor with earth return cannot be considered as a single isolated conductor with any reasonable degree of rigor. Such a circuit would today be treated by reducing it to an equivalent two-conductor problem, for which the inductive effect is more easily and clearly expressible in engineering terms. The two-conductor circuit did not come into practical use, however, until the early 80's when it was found that the ground return introduced too much extraneous noise for satisfactory telephonic communication.

It may be said that the advent of the telephone into the communications field aroused a more concerted effort toward a better understanding of the theoretical behavior of long lines because of the fact that good telephonic communication requires a much more carefully designed transmission circuit than is needed for satisfactory telegraphic communication. This statement must be interpreted, of course, in terms of the speeds which were in use for telegraphic circuits in that early period. The high-speed telegraphic circuits of today require as careful a consideration of details as do any of the other facilities. At all events, this fact partially accounts for the almost thirty-year gap of relative inactivity which is evidenced in connection with the communications aspect of the long-line problem from the time of Kirchhoff's and Weber's work until the more detailed investigations of Oliver Heaviside¹ were made in 1886-87. The only other significant contribution of an earlier date was also made by Heaviside² in 1876 when he formulated the inductance concept in a more rigorous and practical fashion than was done by either Kirchhoff or Weber.

It must also be remembered, of course, that to Kirchhoff and Weber the Maxwellian concept of the electromagnetic field was not yet available as it was to the later investigators. Kirchhoff and Weber did not view the problem from the field standpoint at all. To them the

¹ O. Heaviside, *El. Papers* 2 (1886-87), p. 119 e.s. and p. 307 e.s.

² *Phil. Mag.* Vol. 1, p. 53 (Aug. 1876).

major seat of the phenomenon of propagation of electricity lay in conductor itself rather than in the medium immediately surround it. Maxwell's concise formulation of the field theory enabled a more convenient and accurate introduction of the capacitance inductance parameters as derived quantities in terms of which an approximate representation of the problem could be carried through.

Although the field theory made possible a far better appreciation of the significant factors involved in the long-line problem, at the same time it indicated that the rigorous situation was extremely more involved than had been supposed. A partially rigorous treatment of the single-circular conductor was first made by J. J. Thomson¹ in 1886 and for the concentric cable by him² in 1889. A fuller treatment of the single-conductor problem was given by A. Sommerfeld³ in 1899. The first successful attempt at an exact treatment of two infinitely parallel wires was made by G. Mie⁴ in 1900.

There have been, of course, numerous other contributors to the subject both within and following the period sketched above, those mentioned are perhaps the most important. The conclusion that we can draw from a rather hasty perusal of all this is that the problem of the propagation of electrical energy along a line is, rigorously speaking, a most complex one. None of the above investigations treat the problem fully and rigorously. For the most part an infinitely long line is assumed; the behavior is investigated only in the steady state; the nature or location of the source of energy and the effect of terminal conditions is left out of the picture entirely and finally, the results are interpreted only within certain frequency limits where expansions of otherwise cumbersome expressions are possible. Some of these frequency limits are not even chosen for practical cases. Heaviside⁵ is the only one who considers the behavior of the source as well as the boundary effects both for the initial or transient behavior and for the steady-state condition. He has first, also, to consider the leakage through the insulation, in which the true significance of the inductance parameter may be appreciated. But his calculations consider primarily only one component of the wave out of a large total, namely, that one which carries the transverse energy. His work is a first approximation only as compared with

¹ Math. Soc. Proc. 17, p. 310, London, 1886.

² Roy. Soc. Proc. 46, p. 1, London, 1889. "Recent Researches in Electricity and Magnetism," Oxford, 1893, p. 262 e.s.

³ Wied. Ann. d. Phys. 67, p. 233, 1899.

⁴ Ann. d. Phys. 2, p. 201, 1900.

⁵ O. Heaviside, El. Theory, Vol. 1, art. 54 e.s.; El. Theory, Vol. 2, art. 393.

other, more rigorous treatments. For the engineer, however, this first approximation is usually sufficient. Heaviside's work, therefore, forms the backbone for all subsequent investigations of an engineering nature. The more rigorous attempts serve merely as a guiding background upon which further extensions as well as the degree of approximation of the present status may be based. A few of these matters will now be discussed in more detail.

3. The concept of guided waves. Before Maxwell¹ the physical picture of the propagation of electricity through a long circuit was more or less that which is frequently presented in elementary textbooks, where the hydraulic analogy to an electric circuit is given for purposes of visualization. That is, the seat of the phenomenon was taken to be within the conductor. What occurred outside the conductor could be neither definitely formulated nor described. The electrical energy was thought of as being transmitted through the conductor which, therefore, became of prime importance. In fact, if we accept this point of view altogether, it becomes impossible to conceive of a flow of electrical energy from one point to another without the aid of an intervening conductor of some sort. It has been the writer's experience that many students are quite wedded to this point of view, so much so, in fact, that to them the propagation of energy without wires (wireless transmission) becomes a thing altogether apart from other forms of transmission involving an intervening conducting medium. From this angle, the transmission by radio and that by means of a line become not only different in concept but entirely different in analytic treatment and physical interpretation.

An appreciation of Maxwell's theory of electromagnetic wave propagation brings the so-called wireless and wired forms of transmission under the same roof, so to speak. They merely appear as special cases of the same fundamental phenomenon. This is due to the fact that Maxwell makes the field the primary seat of all that takes place. The presence of a conductor merely causes the field to be broken up into various components, some of which are assigned to the conductor itself, others to the surrounding medium, and still others to the surface separating the two media. The process of analyzing a given situation consists of a simultaneous consideration of the fundamental equations which the field components must obey within the media as well as the various continuity conditions which must be met by the electric and magnetic fields at the surfaces dividing one medium from another.

It is easy to appreciate that to carry out such a line of attack on

¹ The terms "before Maxwell" and "after Maxwell" are used to refer to the periods before and after the concise formulation of Maxwell's field equations.

any but the simplest cases becomes mathematically a huge task, mainly because the method is so rigorous regarding the properties of the various media as well as the exact geometry of the configuration and the conditions at the boundaries. It is for this reason that approximations must be made, even in the so-called rigorous treatment. These approximations usually consist in neglecting certain field components relative to other, much larger ones. The outstanding difference between the usual rigorous treatment and the usual engineering treatment of the problem, however, lies in the fact that in the former the neglected terms are first demonstrated to be negligible, whereas in the latter their existence is often overlooked entirely.

If suitable approximations are made in the rigorous treatment of the usual transmission line with sending equipment at one end and an arbitrary load condition at the other, we find that, for a frequency range including the highest in practice today, only three electromagnetic wave components survive.¹ In their usual order of magnitude, starting with the largest, these are: the transmitted wave, the component supplying line losses, and the radiated wave which represents a loss into the surrounding space.

Of the three, the first, or the transmitted wave, is very much larger than the second, and the second is usually very much larger than the third. This last one, representing the radiated energy, has been shown to be proportional, in the steady harmonically excited state, to the square of the ratio of the distance between conductors to the wavelength, for the fundamental natural frequency of the line,² and to the ratio of the distance between conductors to the length of the impulse.³ In both, the radiated energy for practical lines is negligible as compared to that lost due to dissipation in the line itself, even at the highest frequencies. The fact that the radiation in the harmonic steady state is proportional to the ratio of wire separation to line length means that an infinite line will not radiate under these conditions; i.e., radiation in the steady state is due to

¹ Other components designated by Sommerfeld as "Nebenwellen" (see "Richtungs- und Differenzgleichungen der Physik," F. Vieweg and S., 1927, p. 5; D. Hondros, Ann. d. Phys. 30, 1909, p. 905) are attenuated so rapidly in practical systems that their presence is altogether negligible. Even in dielectric systems "Nebenwellen" have been found experimentally to be extremely small compared to the main wave components (D. Hondros and P. Debye, Ann. d. Phys. 30, 1909, p. 465).

² For harmonics of this fundamental it is proportional to the square of the harmonic number.

³ C. Manneback, "Radiation from Transmission Lines," A.I.E.E. Transactions, p. 289, 1923.

takes place in the vicinity of the line terminals entirely!¹ This argument is hardly applicable to an infinitely long line, however, because there no steady state could ever build up, so that the phenomenon would come under the head of a transient radiation which is independent of the line ends. The result does clearly show why, for finite lengths, the radiation from a transmission line is negligible, whereas for an antenna it is very appreciable.

With the radiation from an ordinary long line negligible, there really remain only two component waves for the engineer to consider, namely, the transmitted and the dissipated waves. The first of these exists in the medium immediately surrounding the conductors, and travels parallel to their axes. Its mode of travel is determined jointly by the properties of the conductors and of the surrounding medium. The line losses are supplied out of the energy in the total wave as it passes along. This constitutes a continuous drain on the transmitted energy. This drain takes the form of a radial electromagnetic wave component which penetrates the surface of the conductor (assumed to have a circular cross-section) and is propagated toward the axis, decaying rapidly as it proceeds inward. The electric field in this component is parallel to the axis of the conductor and thus causes a current to flow in the axial direction. This current is that which is usually designated as the "transmitted" current which eventually "supplies power" to the load at the end of the line. In this formulation of the phenomenon, however, the line current appears merely as an incidental factor which causes a loss to be supplied by the major storehouse of energy residing in the surrounding medium and traveling toward the loaded end. In fact, the conductors themselves become merely the agents whose surfaces guide the surrounding wave to its ultimate goal. In order to emphasize this point of view, we speak of the transmitted wave as the **guided wave**.

The total electromagnetic wave surrounding the conductors has its surface of constant phase slightly tipped toward the conductor. The direction of propagation, which is normal to this surface, is therefore almost axial but with a slight radial inclination towards the conductors. The energy vector, which coincides with this direction, may therefore be split into an axial and a radial component. The former represents the true transmitted energy at the particular point in question; the latter represents the radial drain mentioned above. The fact that this radial component decays as it penetrates the conductor causes the accompanying electric field and hence the conductor current to become weaker nearer the axis, while it is greatest at the surface. This fact is

¹ See also G. Mie, *loc. cit.*, p. 248.

designated as the *skin effect*, which, by means of the above interpretation, becomes a logical and very easily visualized phenomenon.

Concerning the guided wave, the above physical interpretation of the transmission of electromagnetic energy brings the problem of transmission line very much closer to that of wireless transmission. The antenna may really be looked upon as the source. The conductor is the earth, and the dividing surface along which the guided wave travels is the earth's surface. The geometry and the boundary conditions, as well as the properties of the media, are different, but fundamentally the problem is the same as that involving the line.¹

4. The semi-rigorous treatment of the long-line problem. In deriving the differential equations regarding the propagation of electromagnetic waves along

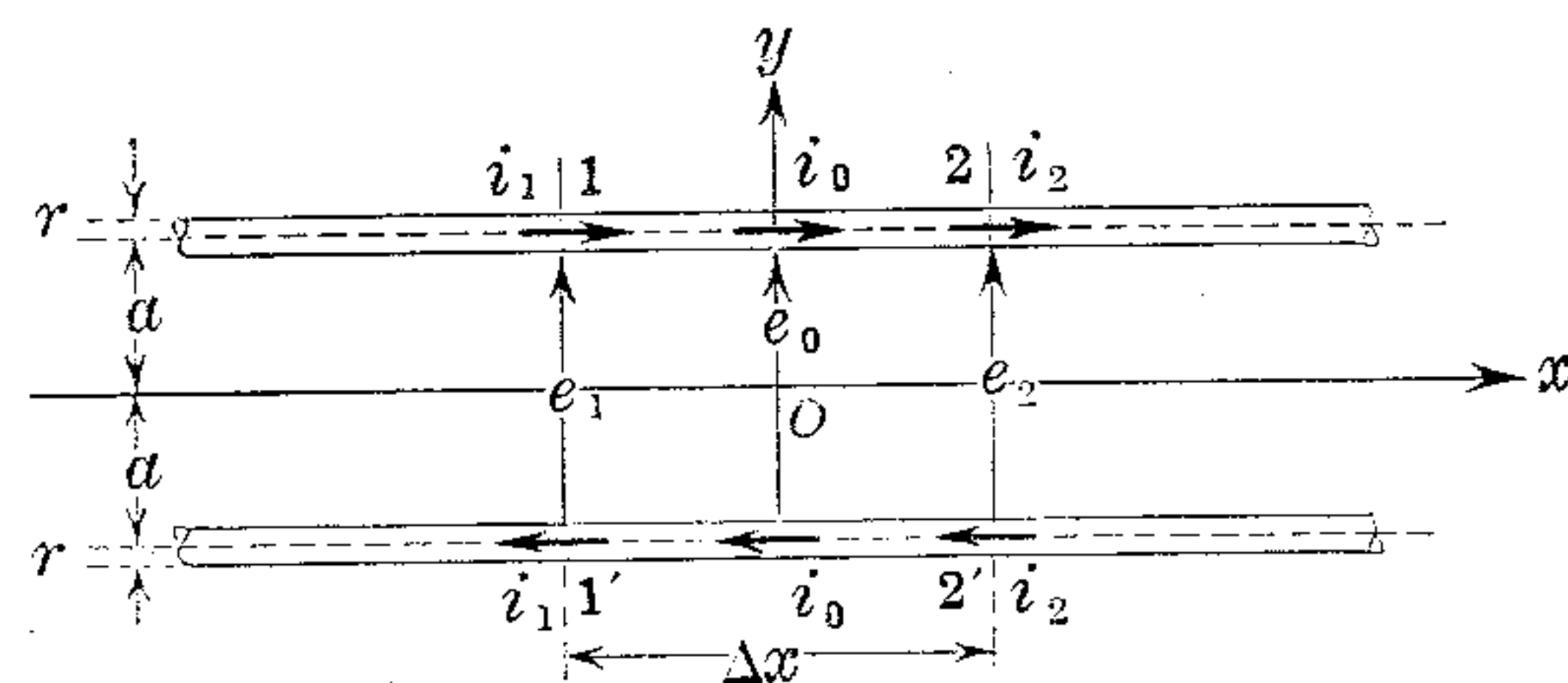


FIG. 1.—Typical internal increment of a parallel two-conductor transmission line.

but more rigorous than the final engineering form involving inductance, resistance, capacitance, and leakage parameters. In other words, our argument proceeds from a middle ground which is neither entirely rigorous nor altogether as approximate and tacit relative to the derivation and significance of the line parameters as the engineering form. This we shall call the *semi-rigorous formulation* of our problem.

The philosophy involved in this formulation is also a middle-ground philosophy, being more like the pre-Maxwellian so far as the circuit relations are concerned, but post-Maxwellian (or “post-Faraday”) in the discussion and interpretation of the electromagnetic field.

To begin with, we shall consider only the presence of the field surrounding the conductors. The fields within the conductors, and their effects upon the distribution of current throughout the conductor cross-section (skin effect), as well as the attendant effect upon the conductor resistances and internal inductances, we shall for the time being neglect.

Fig. 1 represents a typical longitudinal section of a uniform

¹ The ratio of the energy in the guided and radiated waves is, of course, all different, so that the received energy may be due primarily to either one of the components or a combination of the two.

conductor line. The x -axis is laid midway between the conductor axes and parallel to these. The length Δx represents a differential length for which we wish to determine the equilibrium conditions. Since the conductors are assumed identical and symmetrically placed, the currents for a given x will be the same in both except for their directions which will be opposite. The boundaries of Δx are denoted by the cross-sections 1 and 2. The voltage between conductors at the boundary 1 is denoted by e_1 , and that at the boundary 2 by e_2 , while i_1 and i_2 are the corresponding currents at these points.

We now define R as the resistance per unit length of the line (both conductors in series), G as the leakage conductance (from one conductor to the other) per unit length, φ as the flux between the adjacent conductor surfaces (flux linkages not including those within the conductor cross-section) per unit length, and q as the charge per unit length carried by either conductor due to the potential difference between them. In particular, we shall write φ_0 and q_0 for the flux and charge per unit length at the point $x = 0$ where the differential length is located. In general, φ and q are functions of x , and vary from point to point for a given instant of time. Neglecting differentials of the second order, the flux linking the length Δx and the charge per conductor for this length are $\varphi_0 \Delta x$ and $q_0 \Delta x$, respectively. The loop resistance and the leakage conductance for this length are $R \Delta x$ and $G \Delta x$.

Retaining differentials of the first order only, we may, therefore, write the following equations for the voltage and current:

$$\left. \begin{aligned} e_1 - e_2 &= R \cdot \Delta x \cdot i_0 + \frac{\partial \varphi_0}{\partial t} \cdot \Delta x \\ i_1 - i_2 &= G \cdot \Delta x \cdot e_0 + \frac{\partial q_0}{\partial t} \cdot \Delta x \end{aligned} \right\}, \quad (2)$$

where the voltage terms on the left are “rises” and those on the right are “drops” in considering the loop 122'1'. This is the same convention as to signs which is used in lumped networks. The first of these equations states that the voltage variation throughout any differential length of line is partly accounted for by resistance drop and partly by an inductive drop. The second states that the current variation is due in part to a leakage current through the insulation and in part to a charging current (displacement current) due to the electrostatic capacitance between the conductors. The values i_0 and e_0 in these equations refer to the point $x = 0$, i.e., to the location of our differential length. It is not necessary to distinguish between the values at the boundaries 1 and 2 for the terms involving these quantities because this would involve differentials of the second order.

In order to put the relations (2) in differential form, we note by means of a series expansion

$$e = e_0 + \left(\frac{\partial e}{\partial x}\right)_0 \cdot x + \frac{1}{2}\left(\frac{\partial^2 e}{\partial x^2}\right)_0 \cdot x^2 + \frac{1}{6}\left(\frac{\partial^3 e}{\partial x^3}\right)_0 \cdot x^3 + \dots$$

we get for $x = \pm \frac{\Delta x}{2}$, respectively,

$$\left. \begin{aligned} e_2 &= e_0 + \frac{1}{2}\left(\frac{\partial e}{\partial x}\right)_0 \cdot \Delta x + \frac{1}{8}\left(\frac{\partial^2 e}{\partial x^2}\right)_0 \cdot \Delta x^2 + \frac{1}{48}\left(\frac{\partial^3 e}{\partial x^3}\right)_0 \cdot \Delta x^3 + \dots \\ e_1 &= e_0 - \frac{1}{2}\left(\frac{\partial e}{\partial x}\right)_0 \cdot \Delta x + \frac{1}{8}\left(\frac{\partial^2 e}{\partial x^2}\right)_0 \cdot \Delta x^2 - \frac{1}{48}\left(\frac{\partial^3 e}{\partial x^3}\right)_0 \cdot \Delta x^3 + \dots \end{aligned} \right\}$$

Writing similar expressions for current, we can form

$$\left. \begin{aligned} e_2 - e_1 &= \left(\frac{\partial e}{\partial x}\right)_0 \cdot \Delta x + \frac{1}{24}\left(\frac{\partial^3 e}{\partial x^3}\right)_0 \cdot \Delta x^3 + \dots \\ i_2 - i_1 &= \left(\frac{\partial i}{\partial x}\right)_0 \cdot \Delta x + \frac{1}{24}\left(\frac{\partial^3 i}{\partial x^3}\right)_0 \cdot \Delta x^3 + \dots \end{aligned} \right\}$$

Substituting the results (4) into (2) and neglecting third order terms we have after passing to the limit $\Delta x \rightarrow 0$

$$\left. \begin{aligned} \left(\frac{\partial e}{\partial x}\right)_0 + \frac{\partial \varphi_0}{\partial t} + Ri_0 &= 0 \\ \left(\frac{\partial i}{\partial x}\right)_0 + \frac{\partial q_0}{\partial t} + Ge_0 &= 0 \end{aligned} \right\}$$

These are the differential equations for the long line in semi-rigid form. They must, however, be supplemented by two further relations, since, as they stand, they involve four unknowns. For technical purposes we should like to eliminate φ_0 and q_0 so as to leave the voltage and current to be determined. This may be done by expressing φ_0 in terms of the line current and q_0 in terms of the line voltage.

In order to carry this out we will make use of the following well-known relationships in electro- and magneto-statics. These, of course, do not rigidly apply here, but, as we shall see later, their use is justified as a close approximation except at very high frequencies.

The relations which we need here are those for the determination of the magnetic and electric fields surrounding parallel circular conductors

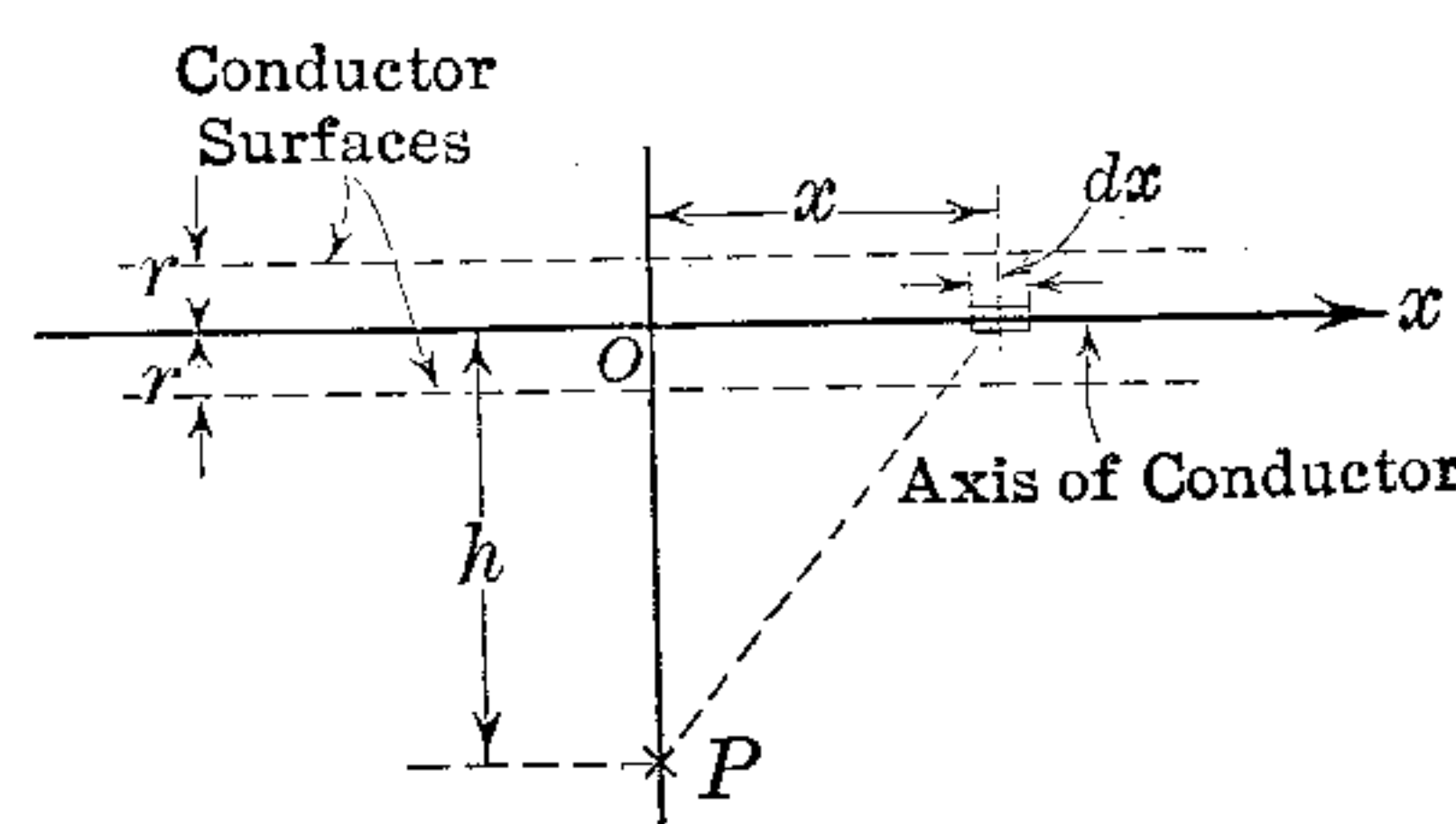


FIG. 2.—Geometrical relationships underlying equations (6) and (7).

carrying prescribed current and charge distributions. For a single linear conductor these are usually derived assuming the current and charge to be concentrated in the conductor axis. Taking the latter as the x -axis, the magnetic and electric field intensities at a point h units perpendicularly from the axis, and located at $x = 0$ (see Fig. 2), are given by:

$$B_p = \mu \int_{-\infty}^{\infty} \frac{h \cdot i(x) \cdot dx}{(h^2 + x^2)^{3/2}} \quad (6)$$

and

$$E_p = \frac{1}{\epsilon} \int_{-\infty}^{\infty} \frac{h \cdot q(x) \cdot dx}{(h^2 + x^2)^{3/2}}, \quad (7)$$

where $i(x)$ and $q(x)$ are functions which give the distribution of current and charge along the conductor. ϵ and μ are the dielectric constant and permeability, respectively, of the surrounding medium. If $q(x)$ is dissymmetrical about $x = 0$, E_p is that component perpendicular to the x -axis. These relations, of course, hold only for $h \geq r$, i.e., for points outside the conductor. The direction of B_p is perpendicular to the plane of the paper and its sense is determined from the direction of the current by the right-hand-screw rule, while that of E_p is in the plane of the paper, perpendicular to the x -axis, and its sense is determined by the sign of the charge q . Other than this, radial symmetry obtains.

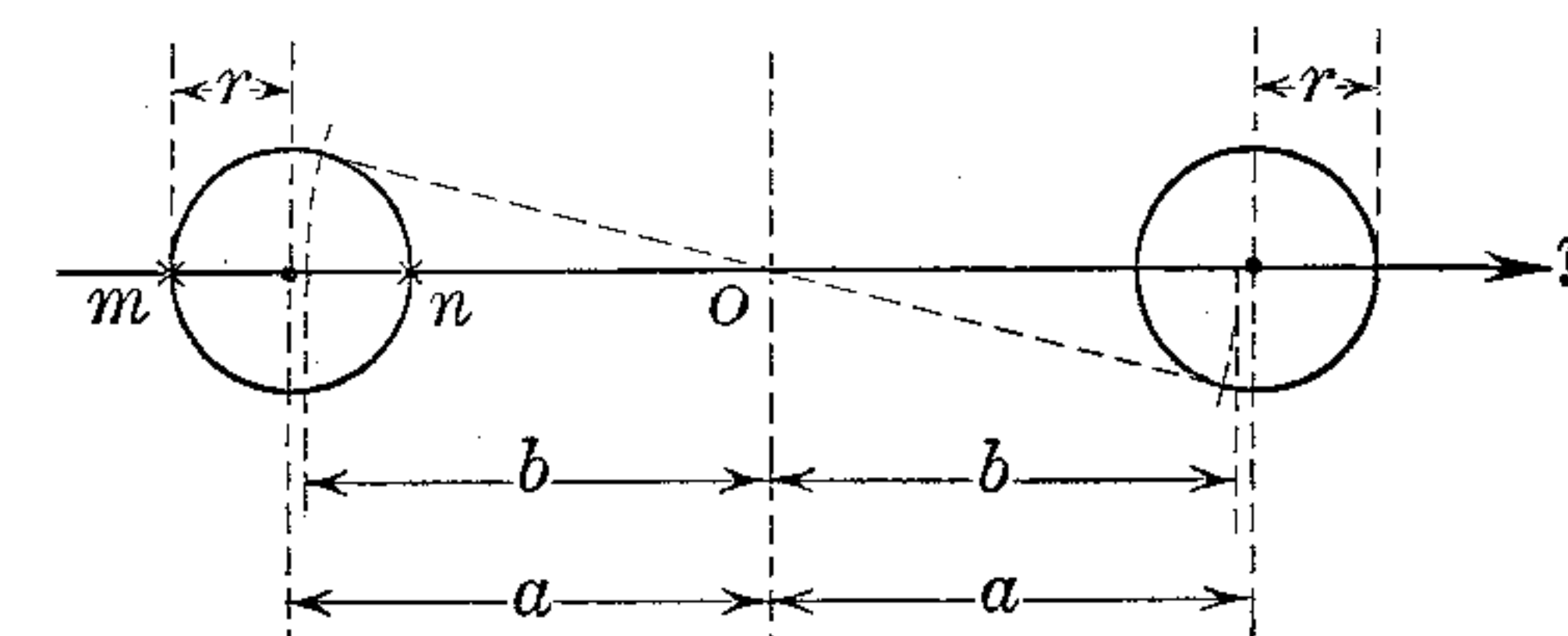


FIG. 3.—Cross-section through the uniform line of Fig. 1.

For a pair of parallel conductors as shown in Fig. 1, the situation is somewhat different. First of all, both conductors contribute toward the values of the resulting fields. In addition to this the nearness of one conductor to the other (proximity effect) also modifies matters. This modification is in general quite complex for the non-stationary state. For stationary currents and charges, however, with the assumption that the resistivity of the conductor material is negligible, the modification merely amounts to the fact that the currents and charges can no longer be considered concentrated along the conductor axes, even for the external field distribution. They can, however, be considered concentrated along axes which are eccentric with respect to the conductor axes and shifted toward the adjacent surfaces of the two

conductors.¹ This is illustrated in Fig. 3, which represents a cross section through the line of Fig. 1. The origin of coordinates is at the center of the figure, and the x -axis is directed perpendicularly into the paper. The conductor on the left, therefore, is the lower one of Fig. 1, and the right-hand one, the upper. The conductor axes are the lines $y = \pm a$, while those axes along which the currents and charges may be considered concentrated are located by $y = \pm b$. As indicated in the figure, the distance b is determined from:

$$b^2 = a^2 - r^2.$$

This may be demonstrated by making use of the fact that the conductor surfaces must be equipotential loci. Since, for a uniform charge distribution along the axes $y = \pm b$, the potential is logarithmic, we get for the potential at any point in the cross-sectional plane, assuming equal and opposite charges, the difference of two logarithms or the logarithm of the ratio of the distances from this point to the axes $y = \pm b$. Constant potential loci, therefore, are such for which this ratio equals a constant. Considering two points m and n (Fig. 3) on the left-hand conductor surfaces, for example, we must have:

$$\frac{a - b + r}{a + b + r} = \frac{-a + b + r}{a + b - r},$$

from which (8) follows.

With this point in mind, we can now apply the relations (6) and (7) to the determination of the magnetic and electric field intensities at a point on the y -axis of Fig. 1, located between the adjacent surfaces of the conductors. Assuming that the top conductor carries the positive charge and the lower an equal negative one, a positive voltage will be a rise from the lower to the upper, and applying the proper rules for the directions of the electric and magnetic field intensities, we see that the vector product, which indicates the positive direction of energy flow, coincides with the positive x -direction. Making this assumption:

¹ So far as the charges are concerned, this may be seen from the fact that the opposite charges on the two conductors attract each other. The amount of the shift is determined by utilizing the fact that the conductor surfaces must be equipotential loci. Why the axes in which the currents may be considered concentrated are shifted, and in fact are coincident with those for the charges, is not so evident. It follows, however, if the conductivity of the wires is assumed infinite. Then the skin and proximity effects, which tend to crowd the current toward the adjacent surfaces, are complete even at the lowest frequency, and may hence be assumed to be so in the limit in the case of zero frequency. This dissipationless or ideal case we shall take as our starting point for our present discussion of the external fields, and later consider the deviations from it.

using the relations illustrated in Fig. 3, the application of (6) and (7) gives:

$$B = \mu \int_{-\infty}^{\infty} \left\{ \frac{b - y}{[(b - y)^2 + x^2]^{3/2}} + \frac{b + y}{[(b + y)^2 + x^2]^{3/2}} \right\} \cdot i(x) \cdot dx \quad (9)$$

and

$$E = -\frac{1}{\epsilon} \int_{-\infty}^{\infty} \left\{ \frac{b - y}{[(b - y)^2 + x^2]^{3/2}} + \frac{b + y}{[(b + y)^2 + x^2]^{3/2}} \right\} \cdot q(x) \cdot dx \quad (10)$$

where the direction of B is perpendicularly into the paper, and that of E is in the y -axis from top to bottom.

The flux per unit length and the voltage between conductors may then be found from:

$$\varphi_0 = \int_{-(a-r)}^{a-r} B \cdot dy \quad (11)$$

and

$$e_0 = - \int_{-(a-r)}^{a-r} E \cdot dy \quad (12)$$

respectively. Substituting (9) and (10) into (11) and (12) and carrying out the integration with respect to y , we find:

$$\varphi_0 = 2\mu \int_{-\infty}^{\infty} \left\{ \frac{1}{[(b - a + r)^2 + x^2]^{1/2}} - \frac{1}{[(b + a - r)^2 + x^2]^{1/2}} \right\} \cdot i(x) \cdot dx, \quad (13)$$

and

$$e_0 = \frac{2}{\epsilon} \int_{-\infty}^{\infty} \left\{ \frac{1}{[(b - a + r)^2 + x^2]^{1/2}} - \frac{1}{[(b + a - r)^2 + x^2]^{1/2}} \right\} \cdot q(x) \cdot dx. \quad (14)$$

These relations, together with (5), completely specify the behavior of the guided wave. If the system were dissipationless, i.e., if R and G were zero, this formulation for the behavior of the guided wave would be exact, because the field distributions upon which the derivations of (13) and (14) depend are rigorously correct at moderate frequencies only under these circumstances.¹ Otherwise not only the basis for (13) and (14) but also the values of R and G will depend upon the frequency.

These relations are, however, in such a form that it would be very difficult if not impossible to determine a solution except by a method of successive approximations. This is due primarily to the fact that

¹ Here it should also be noted that (13) assumes the flux to be due to the conductor current alone. At very high frequencies the displacement current may contribute sufficiently to the magnetic field intensity to invalidate this result. For frequencies satisfying the criterion set up in the next section, however, the contribution due to the displacement current is negligibly small. Granting this for the moment, we may accept (13) as sufficiently representing the situation for our purposes.

the functions $i(x)$ and $q(x)$ are involved in the integrals (13) and (14). The engineering statement of the problem amounts to forming a first approximation toward the solution of (5), (13), and (14).

5. The engineering treatment of the long-line problem. In order to evaluate the integrals (13) and (14) approximately, we may assume that

$$\left. \begin{aligned} i(x) &= i_0 \\ q(x) &= q_0 \end{aligned} \right\} \quad (16)$$

i.e., that these functions have everywhere the same value as at point $x = 0$, which refers to the arbitrary location of Δx in Fig. 1. This is the same as assuming that the variations of $i(x)$ and $q(x)$ in (13) and (14) affect the values of these integrals to a negligible extent. This in turn will be the case if the rates of variation of $i(x)$ and $q(x)$ are small compared to the rate of convergence of the integrals assuming $i(x)$ and $q(x)$ to be constant.

In order to investigate this, we will replace the limits in these integrals by finite values, and evaluate them under the assumptions (16). The rate of convergence of the integrals can then be studied by considering this result as a function of the finite limits. To carry this out we form the integral

$$\begin{aligned} J &= \int_{-k}^k \left[\frac{1}{[(b-a+r)^2 + x^2]^{1/2}} - \frac{1}{[(b+a-r)^2 + x^2]^{1/2}} \right] \cdot dx \\ &= 2 \ln \left(\frac{b+a-r}{b-a+r} \right) - 2 \ln \left(\frac{k + \sqrt{k^2 + (b+a-r)^2}}{k + \sqrt{k^2 + (b-a+r)^2}} \right), \end{aligned}$$

which is the integral common to both (13) and (14) after $i(x)$ and $q(x)$ are taken out. Using the relation (8) and retaining only the first significant term in an expansion of the second logarithm,¹ we find:

$$J = 2 \ln \left(\frac{a + \sqrt{a^2 - r^2}}{r} \right) - \frac{2(a-r)\sqrt{a^2 - r^2}}{k^2}. \quad (17)$$

Here the first term represents the value which the integral would have with infinite limits. The second term represents the approximate error made by replacing the infinite by finite limits. Writing this as

$$J = 2 \ln \left(\frac{a + \sqrt{a^2 - r^2}}{r} \right) \cdot (1 - \epsilon),$$

¹ The radicals are first expanded in terms of $[(b+a-r)/k]^2$ and $[(b-a+r)/k]^2$. Dropping terms above the square, the argument is simplified to $\{1 + b(a-r)/k^2\}$. Since $b(a-r) \ll k^2$, the logarithm of this quantity except for higher order terms is given by $b(a-r)/k^2$.

we find the decimal error to be:

$$\epsilon = \frac{(a-r)\sqrt{a^2 - r^2}}{k^2 \ln \left(\frac{a + \sqrt{a^2 - r^2}}{r} \right)}. \quad (17)$$

Noting that $a > r$, we see that, for a given k , this error becomes smaller, and hence the rate of convergence of the integral greater, for a smaller ratio of line spacing to wire diameter. For a given spacing and wire diameter, the error decreases inversely as k^2 . For a prescribed error, on the other hand, the value of k may easily be determined. In terms of the latter, we may now formulate a criterion regarding the degree of approximation involved in assuming $i(x)$ and $q(x)$ constant for the evaluation of (13) and (14), namely: *If $i(x)$ and $q(x)$ vary so slowly as to be sensibly constant throughout the interval $-k < x < k$, then the error involved in assuming them constant throughout the interval $-\infty < x < \infty$ will not exceed the approximate value given by (17).*

In order to illustrate this numerically, suppose we consider an open telephone line with No. 10 conductors spaced 10 inches on centers. Assuming an allowable error of 5 per cent, we find by (17) that k is roughly 10 inches. Hence if the variations of current and voltage along the line (for a given instant) are slow enough so that they may be considered constant throughout a 20-inch length, then for this line the above approximation may be made with sufficient accuracy for engineering purposes.

The question as to whether $i(x)$ and $q(x)$ may be considered slowly variable depends, in the harmonically excited steady state, upon the frequency, or rather upon the wave length as determined from (1). We shall see later that, for neglected losses in the system,¹ the current and charge are simple harmonic functions of x . A wave length is that distance throughout which the functions complete one cycle. Throughout a small portion of a wave length they may be considered fairly constant. That is, if the wave length is large compared to $2k$, then the above criterion will be approximately met.

This situation is actually more favorable than the above argument would indicate, for the following reason. Any actual variations of current and charge with distance can always be approximated by the following series expansions:

$$\left. \begin{aligned} i(x) &= i_0 + i_1 \cdot x + i_2 \cdot x^2 + \dots \\ q(x) &= q_0 + q_1 \cdot x + q_2 \cdot x^2 + \dots \end{aligned} \right\} \quad (18)$$

¹ With a reasonable amount of dissipation present, the situation is usually only slightly modified.

The assumptions (15) restrict these expansions to their first term only. Now we note that the integrand in (13) and (14), except for the factors $i(x)$ and $q(x)$, is an even function of x . If (18) is substituted, and the integration carried out term by term, all terms for the odd powers in (18) vanish because their integrands are odd functions of x and hence the integrals between symmetrical limits are zero. This means that the odd terms in (18) have no effect upon the values of (13) and (14)! Consequently the current and charge may vary linearly (for example) throughout the interval $-k < x < k$ at any rate at all, and yet the result will be the same as though they were constant. The only kind of variation within this interval which disturbs the value of the integrals is a symmetrical one.

For our simple harmonic variation of $i(x)$ and $q(x)$ in the dissipationless case, this means that the vicinity of a node¹ will give rise to the same value of the integral (13) or (14) as would a constant region with the same average value. An anti-node,² on the other hand, is a symmetrical region. This would influence the values of the integrals. But the variation in the vicinity of an anti-node is rather slow. The cosine function, for example, varies about an average value by only 10 per cent from $-\frac{\pi}{4}$ to $+\frac{\pi}{4}$, i.e., over a quarter of a wave length in the vicinity of its maximum.

From this we may conclude that if one-quarter wave length is equal to or larger than $2k$ as determined from (17) with a reasonable value for ϵ , then the assumption as to the constancy of $i(x)$ and $q(x)$ in (13) and (14) may be made with sufficient accuracy for most engineering purposes.³ Fixing the allowable value of ϵ at five-hundredths, then

¹ A point where the function passes through zero.

² A point where the function passes through its maximum.

³ For frequencies in this vicinity or higher, the displacement current begins to have an effect upon the magnetic field intensity which determines φ_0 . If we assume a non-dissipative system, equations (68a) and (69a), p. 56, next chapter, show that the displacement current, which depends upon the time derivative of the voltage, is in time phase and in space quadrature with the conduction current. Both vary sinusoidally with distance. The displacement current has a node where the conduction current has an anti-node, and vice versa. Thus, in the vicinity of a conduction current node, the displacement current is a symmetrical (even) function and hence contributes nothing to φ_0 , whereas in the vicinity of an anti-node of the conduction current the displacement current is an anti-symmetrical (odd) function so that the error in φ_0 made by using i_0 for i seems partially to be compensated for by a contribution from the displacement current. Although this argument would extend the validity of the engineering assumption beyond the criterion (19), a determination of the effect of displacement current at these high frequencies requires treatment which accounts for retardation. In this light the compensation is not

means that the engineering statement of the case, which assumes (15), is roughly justified so long as:

$$\lambda \geq 36 \sqrt{\frac{(a-r)\sqrt{a^2-r^2}}{\ln\left(\frac{a+\sqrt{a^2-r^2}}{r}\right)}} \quad (19)$$

where λ denotes the wave length. The equality sign gives the shortest wave length for which reasonable results may be expected from an engineering analysis. For the open line with No. 10 conductors and 10-inch spacing, this gives a smallest wave length of 78 inches. Assuming the phase velocity to be that of light, the relation (1) gives the corresponding highest frequency as $152 \cdot 10^6$ cycles per second, which is many times the highest essential frequency ever occurring on such a facility.

This argument is based upon the harmonically excited steady state. The situation is quite otherwise for transient conditions involving steep wave fronts. In the vicinity of the wave front, the above procedure with regard to the evaluation of (13) and (14) is not at all justified. It is quite obviously also not justified at points closer than k units to the line ends. How the engineering procedure is to be justified in such cases is not at all clear.¹ The fact that experimental results substantiate the engineering solution fairly well, even in the transient cases, seems to indicate that, where the criterion (19) is exceeded, errors of a compensating nature appear. It is also very likely that wave fronts so steep as to invalidate the assumption regarding reasonably constant values of i and q over the region $-k < x < k$, do not occur in any transient disturbances encountered in practice.

what it seems to be according to the above argument. For our purposes it is unnecessary to enter more deeply into this discussion.

¹ In an article entitled "The Guided and Radiated Energy in Wire Transmission" (A.I.E.E. J., Oct., 1924, p. 908), J. R. Carson points out that for the steady-state behavior the effect of the line ends is to introduce an infinite number of complementary waves which are ordinarily very rapidly attenuated toward the interior of the line and are, therefore, entirely negligible in comparison with those given by the usual engineering solution. No mention is made, however, as to the legitimacy of expressing these solutions in terms of inductance and capacitance parameters or the manner in which such a procedure depends upon the frequency and geometry involved. In a subsequent paper ("The Rigorous and Approximate Theories of Electrical Transmission Along Wires," B.S.T.J., Vol. 7, pp. 11-25, Jan., 1928), Carson takes up a discussion of essentially the same questions raised in the treatment given here, with no further elucidation of the effect of physical boundaries except the mention that their consideration leads to an extremely involved analysis.

The criterion (19) takes on an interesting form if we divide through by r and introduce for the ratio of wire separation to wire diameter

$$s = \frac{a}{r}.$$

Then we have

$$\frac{\lambda}{r} \geq 36 \sqrt{\frac{(s-1)\sqrt{s^2-1}}{\ln(s + \sqrt{s^2-1})}}, \quad (19a)$$

or, dividing through by s ,

$$\frac{\lambda}{a} \geq \frac{36}{2} \sqrt{\frac{(s-1)\sqrt{s^2-1}}{\ln(s + \sqrt{s^2-1})}}. \quad (19b)$$

These forms express the shortest wave length in terms of either the wire separation or the wire size for a given ratio between these quantities. In the above example $s = 100$, which gives

$$\frac{\lambda}{r} \geq 1556; \frac{\lambda}{a} \geq 15.56.$$

For a given ratio and wave length, the criterion imposes an upper limit upon the wire size or the spacing.

Recalling what was said earlier regarding lumped and distributed systems, the above formulation may be interpreted as characterizing the long line as a lumped system so far as its transverse dimensions are concerned, but as a distributed system in the longitudinal direction. This is evidenced by the criterion (19) which compares only the transverse dimensions with the wave length. The engineering formulation of the long-line problem is thus a peculiar hybrid combination with respect to lumped and distributed characteristics.

This fact makes it necessary to impose a restriction upon the definitions of voltage and current as they normally appear under electrostatic and magneto-static conditions. The voltage must be defined as the line integral of electric field intensity taken from a point on one conductor to a point on the other conductor *in the same cross-section plane*. Here the path of integration is restricted to lie wholly in the plane, but may otherwise be arbitrary.¹ This restriction must be imposed for the reason that the electric field may be considered conservative only in the two transverse dimensions. Longitudinally the field is non-conservative, and an arbitrary path which utilizes the third dimension can lead to a different value for the line voltage. So long as the criterion (19) is fulfilled, however, the electric field is substantially conservative in the transverse dimensions, so that the line voltage may still be defined in this restricted sense. A similar situation holds with regard to the line current. The latter, according to Ampère's circuital law, is given by the line integral of magnetic field intensity encircling either conductor. This integral also will not be independent of the path unless the path is restricted to lie wholly in the cross-sectional plane and unless the criterion (19) is met. The concepts of current and voltage, in terms of which the engineer prefers to characterize all network phenomena, may therefore be retained in this restricted sense for describing long-line behavior in all practical cases except those involving ultra short waves for which the criterion (19) is seriously violated.

¹ This assumes an infinitely long line. If the line is finite the field at points far from the conductors may be materially influenced by the radiation components due to the end effects. The present argument leaves the latter out of consideration.

Having thus substantiated the engineering procedure, we will continue with the formulation of the equilibrium conditions in more familiar terms. Neglecting the second term in (16), we find for (13) and (14)

$$\varphi_0 = 4\mu i_0 \ln\left(\frac{a + \sqrt{a^2 - r^2}}{r}\right), \quad (20)$$

and

$$e_0 = \frac{4q_0}{\epsilon} \ln\left(\frac{a + \sqrt{a^2 - r^2}}{r}\right). \quad (21)$$

From these we get

$$\frac{\partial \varphi_0}{\partial t} = \frac{\partial i_0}{\partial t} \cdot 4\mu \ln\left(\frac{a + \sqrt{a^2 - r^2}}{r}\right), \quad (22)$$

and

$$\frac{\partial q_0}{\partial t} = \frac{\partial e_0}{\partial t} \cdot \frac{\epsilon}{4 \ln\left(\frac{a + \sqrt{a^2 - r^2}}{r}\right)}. \quad (23)$$

If we now let

$$L = 4\mu \ln\left(\frac{a + \sqrt{a^2 - r^2}}{r}\right), \quad (24)$$

and

$$C = \frac{\epsilon}{4 \ln\left(\frac{a + \sqrt{a^2 - r^2}}{r}\right)}, \quad (25)$$

then substitution of (22) and (23) into (5) gives

$$\left. \begin{aligned} \frac{\partial e_0}{\partial x} + L \frac{\partial i_0}{\partial t} + R i_0 &= 0 \\ \frac{\partial i_0}{\partial x} + C \frac{\partial e_0}{\partial t} + G e_0 &= 0 \end{aligned} \right\}. \quad (26)$$

These relations involve only the current and voltage, and are the desired equilibrium conditions for engineering purposes.

The quantities (24) and (25) are the distributed inductance and capacitance parameters per unit length. The product of these two

$$LC = \epsilon\mu, \quad (2)$$

and thus depends only upon the properties of the medium surrounding the conductors. This is to be expected since only the external field was considered in the derivation of L and C . However, the fact that the product of L and C is independent of the geometry of the system is a rather striking result. In this connection it should be remembered that the formulae (24) and (25) are rigorously correct only in the dissipationless case, i.e., for conductor material having zero resistivity and a dielectric having zero conductivity. In the dissipative case they are very nearly correct in the vicinity of those extremely high frequencies which approach the limit indicated by (19). These matters will be discussed more fully below.

In connection with the relation (27) it should be remembered that for free space the product of ϵ and μ in any cgs system of units¹ equals $1/9 \cdot 10^{20}$ and has the dimensions of the reciprocal of the square of velocity. In the practical cgs system, for example, $\epsilon = 1/9 \cdot 10^{11}$ and $\mu = 10^{-9}$. With these values the relations (24) and (25) come out in practical henries per centimeter and practical farads per centimeter respectively. For free space, therefore,

$$\frac{1}{\sqrt{LC}} = 3 \cdot 10^{10} \text{ cm per sec,}$$

which is the velocity of light. We shall see later that this is also the velocity of phase propagation of electromagnetic waves along a dissipationless open line.

6. Discussion of the line parameters. From the theoretical as well as the practical standpoint, it is important to bear in mind several matters regarding the proper determination of values for the line parameters R , L , G , and C . Since the practical considerations involved in this discussion are in some instances different from the theoretical, we shall first take up the latter, and then separately point out how they are modified by circumstances which are usually present in practice.

First we wish to point out again that the above derivations of L and C are correct only for a dissipationless system. For such a system

¹ This refers to the electrostatic, electromagnetic, or practical systems only. In the Gaussian or mixed system of units, both ϵ and μ are unity, but there a factor (the square of the velocity of light) enters into the derivation so as to give rise to just the same.

the field distributions are entirely outside of the conductors at all frequencies, and are correctly given by considering the charges and currents concentrated along the axes $y = \pm b$ as shown in Fig. 3. This is true up to frequencies at which the displacement current parallel to the conductor axes becomes appreciable, beyond which the engineering formulation can no longer be used without modification. The dissipationless case is therefore the ideal one in which the bothersome variation of the line parameters with frequency is entirely absent.

As soon as the resistivity of the conductor material is considered, the situation becomes markedly different, particularly with respect to the inductance parameter. The capacitance parameter varies so little with frequency, even with an appreciable amount of conductor resistance, that it may be considered constant and given by (25) throughout the practical frequency range. This is due primarily to the fact that the charges may still be considered concentrated along the axes $y = \pm b$ in Fig. 3, even when resistance is present, so that the electric field distribution remains unchanged and hence the same procedure in the derivation of C applies. This is not so in the case of the inductance. As soon as the slightest amount of resistance is present, the skin and proximity effects for zero frequency, or for frequencies very close to zero, are entirely absent instead of being substantially complete. That is to say, the current distribution throughout the conductor cross-sections is uniform for zero frequency as well as almost uniform for very small frequencies. Hence the conductor currents must be considered concentrated along the axes $y = \pm a$ instead of along $y = \pm b$ so far as the external magnetic field distribution is concerned. This means that in the derivation of the inductance parameter we must replace b by a in the integral (9). If we do this, we obtain instead of (24)

$$L = 4\mu \ln\left(\frac{2a - r}{r}\right). \quad (24a)$$

This is the correct value for the inductance due to the external field in the stationary state (zero frequency) for the dissipative case. Here the current is uniformly distributed throughout the conductor cross-sections, so that the magnetic field penetrates into the interior of the conductors instead of being altogether external as in the dissipationless case. Hence (24a) must be supplemented by that contribution to the total inductance due to the internal flux linkages. This amount is found to be¹

$$L_i = \mu_i \quad (24b)$$

¹ See for example L. F. Woodruff, "Principles of Electric Power Transmission and Distribution," John Wiley & Sons, 1925, p. 17, e.s.

for the stationary state, μ_i being the permeability of the conductor material. The total inductance for zero frequency is the sum of (24a) and (24b) as soon as any amount of resistance is present.

As the frequency is increased gradually from zero to the upper limit discussed above, the value of the total inductance parameter varies continuously from its constant-current value to the value given by (24). The latter, which is the dissipationless value at all frequencies, is correct for the general dissipative case only at the high-frequency limit where skin and proximity effects are complete so that all the flux linkages are external, and the external field distribution is given by the picture which was had for the non-dissipative case. This means that, as the frequency increases from zero, the internal portion (24b) gradually disappears, and the external portion (24a) gradually merges with the value (24). If the conductor resistance is small, this variation proceeds very rapidly as the frequency increases. In fact, if the resistance is considered extremely small, the transition from the sum of (24a) and (24b) to the value (24) is practically complete even for very small frequencies. In the case of vanishing conductor resistance, it is complete from the start. This is the line of reasoning which justifies (24) in the dissipationless case even for zero frequency, which is there looked upon as a vanishingly small frequency.

Since the value (24a) is smaller than (24), we see that the external inductance increases with frequency, the percentage increase depending upon the geometry of the configuration, i.e., upon the distance a compared to r . The internal inductance decreases as the frequency increases. How the net inductance varies throughout the frequency range is quite a complex problem, and depends upon the geometry. If the spacing of the conductors is large compared to their radii, the difference between (24a) and (24) is inappreciable, and the net change is almost wholly attributable to that of (24b), which is a decrease with frequency. If the conductors are closely spaced, as in cables, (24) may be considerably larger than (24a), and the net change may be an increase.

The exact theoretical variation of the resistance parameter with frequency is due to changes in the magnitude and distribution of internal flux linkages. The net effect is the result of two tendencies, one of which is to crowd the current toward the surface in all radial directions, and the other to increase the current density in the adjacent conductor surfaces. Although the first of these (the skin effect) tends toward an infinite surface density as the frequency becomes infinite, the second (proximity effect) merely tends toward a maximum or uniform circumferential distribution which still leaves some current at

points on the conductor contour. The skin effect, of course, increases the conductor resistance materially with frequency, and tends to make it infinite for an infinite frequency. The superposed proximity effect, which becomes more pronounced for a smaller ratio of separation to conductor radius, may greatly increase the resistance over that due to the normal skin effect at a given frequency. In open telephone lines, the ratio a/r is large, and the proximity effect almost negligible. In cables, however, it may cause the resistance to increase with frequency many times faster than it would from the skin effect alone.

Finally the variation of the leakage parameter with frequency may also be theoretically determined on the assumption of a constant conductivity of the dielectric, which is supposed to surround the conductors uniformly and extend indefinitely in the transverse directions. Under these conditions, the change in the leakage parameter is due solely to changes in the external field distributions which, as we have seen, are very slight for all frequencies up to the limit (19). The calculated variation in the leakage parameter is thus negligible.

Theoretically, therefore, the parameters G and C may be considered constant for all frequencies for which the engineering formulation of our problem is valid. Of the parameters R and L , the first is the more variable, particularly where the ratio a/r is small. Starting with its d-c value, this parameter tends toward infinity as the frequency increases without limit. The inductance may increase or decrease with frequency depending upon the ratio a/r . In any event the variation tends toward the dissipationless value at the upper frequency limit, and hence the variation is never comparable with that of R . For large a/r ratios the variation in the total inductance is due almost entirely to the internal portion, which in such cases is a small fraction of the total, so that the entire variation may be negligible.

In actual telephone lines and cables, the variation of R with frequency is found to agree, within experimental error, with the theoretically predicted variation.¹ This variation in the usual open telephone lines employing copper conductors may run from 250 to 400 per cent over a frequency range of 500 to 50,000 cycles, the larger variation occurring for the larger conductor sizes. Owing to the relatively large a/r ratio, this variation in open lines is due almost entirely to skin effect. In lines employing iron wire, the variation is considerably larger because of the higher internal permeability. In cables the *percentage* variation in R , although contributed to by the proximity effect, is considerably

¹ E. I. Green, "Transmission Characteristics of Open-Wire Telephone Lines," B.S.T.J., Oct., 1930.

less on account of the smaller wire size and consequent higher inductance.

The variations in the parameters L and C for practical circuits agree substantially with theoretical predictions. For open lines inductance decreases roughly from 2 to 4 per cent over a frequency range from zero to 50,000 cycles, the larger variation occurring for larger conductor sizes and smaller spacings. The capacitance parameter which theoretically should not vary by any measurable amount, does increase from 0.5 to 4 per cent over the same frequency range in open wire lines. This is due to the fact that the theory does not take into account the effect of insulators, pins, and cross-arms which are necessary in the support of open lines. The use of metal pins, or pins with metal thimbles, introduces additional lumped capacitance at the support points, particularly when pairs of pins or thimbles are bonded. When numerous leakage paths over the surfaces as well as through the wires of the insulators are taken into account, these lumped capacitances may be approximately attributed to the resulting capacitance effect of a ladder network consisting of small condensers and large leakage resistances.¹ This rather complicated structure presents a capacitive component which has a rising frequency characteristic. It is this that accounts for the small observed increase in the net capacitance parameter on such facilities.

The widest difference between theoretical and observed variations is found in the leakage parameter. This, however, is to be expected since the theoretically assumed constancy of the dielectric conductivity is quite far from the truth. Whereas the theory predicts no observable variation in the leakage conductance, it is found both in open telephone lines and in cables to be *almost a linear function of frequency* over a range of 1000 to 50,000 cycles.² For open lines, in dry weather, the variation is roughly 5000 per cent of the low-frequency value, and in cables it is more than 8000 per cent! In addition to this, the dry weather values for open lines are from three to twelve times the dew weather values, the larger differences being observed at the low frequencies.

It may seem to the reader that, in view of such variations in the line parameters, the long transmission line could hardly be treated analytically according to methods which rigorously apply only to constant-parameter networks. The variation in the leakage conductance seems particularly discouraging from this point of view. But, we shall see later, the effect of the leakage parameter upon the per-

¹ L. T. Wilson, "A Study of Telephone Line Insulators," B.S.T.J., Oct., 1930.

² E. I. Green, *loc. cit.*

formance of open lines and cables is slight if not altogether negligible over the usual frequency ranges for which such facilities are used. The resistance, which is the only other parameter whose variation is appreciable, may alone be considered as seriously affecting this question. At a given frequency, however, the parameters are constant. The solution for a single frequency will, therefore, be correct provided the proper parameter values for that frequency are used. Obviously the superposition of a finite number of frequencies may be treated in this manner also. A real difficulty arises in the treatment of the transient behavior, for which average values of the parameters will have to be selected unless this situation is considered in the Fourier sense, whereby the problem is again expressed in terms of the harmonic steady state. These matters, however, will be more fully appreciated later on. In the meantime it is quite evident that a treatment which assumes the constancy of the line parameters, in spite of what actual cases may seem to indicate, will have to be the point of departure in all subsequent modifications. The following chapters, therefore, are devoted to a thorough study of such a theoretically ideal line.

7. Bibliography. In the above discussion our object has been twofold. First, to acquaint the reader somewhat with the basis upon which the engineering considerations of the long-line problem rest so that he will have a better appreciation of their limitations; and second, to arouse his interest in pursuing this problem from a more rigorous angle at a later date when he shall have become thoroughly familiar with the elementary treatment. Such a further study in the region beyond which the present formulation may be applied certainly seems justified in view of the increasing tendency toward the use of higher frequencies in communication work, particularly in the field of radio. This applies equally to networks which at present are considered lumped as it does to uniform systems of larger dimensions.¹ Such further study will, of course, require a more thorough knowledge of the fundamental concepts involved than may be gained from the present rather brief survey; hence the literature, both early and recent, which bears upon this subject will have to be consulted. Rather than append a lengthy bibliography here, we refer the reader for an earlier survey (up to 1906) to a summary by M. Abraham² where many additional references will be found. For a more recent summary with bibliographic notes, the article

¹ For a general discussion of this situation see J. R. Carson's paper, entitled "Electromagnetic Theory and the Foundations of Electric Circuit Theory," B.S.T.J., Vol. 6, pp. 1-17, January, 1927.

² M. Abraham, "Elektromagnetische Wellen," Encyclopädie der mathematischen Wissenschaften, Vol. V², pp. 514-538.

by C. Manneback already referred to will be found useful. A rather comprehensive treatment, particularly with regard to the theoretical variation of line parameters with frequency, will be found in a Bureau of Standards publication by C. Snow.¹

¹ Chester Snow, "Alternating Current Distribution in Cylindrical Conductors," Scientific Papers of the Bureau of Standards, Vol. 20, pp. 277-338, 1925; or page S 509.

PROBLEMS TO CHAPTER I

1-1. A two-wire transmission line which is to be used as a feeder for an antenna has a ratio of wire separation to wire diameter of 100. The resistance and leakage may be neglected.

(a) What are the distributed inductance and capacitance parameters in practical henries and farads per meter?

(b) If a frequency corresponding to a wave length of 30 meters is to be transmitted, what is the maximum wire separation for which an engineering analysis may be expected to be reasonably correct?

1-2. Consider a two-wire transmission line whose conductor axes are spaced d centimeters apart and whose common radius is r centimeters with $r \ll d$. The conductors are assumed to carry uniform charge distributions of q_1 and q_2 coulombs per unit length. Show that the respective conductor potentials e_1 and e_2 may be expressed in terms of the charges by means of the linear relations

$$\begin{aligned} e_1 &= s_{11} q_1 + s_{12} q_2 \\ e_2 &= s_{21} q_1 + s_{22} q_2 \end{aligned}$$

in which the coefficients (called the partial elastance coefficients) are given by

$$s_{11} = s_{22} = \frac{2}{\epsilon} \ln \frac{1}{r}; \quad s_{12} = s_{21} = \frac{2}{\epsilon} \ln \frac{1}{d}.$$

These logarithmic potentials are understood to be determined to within a common additive constant and are referred to a common datum located at a distance X centimeters from one of the conductors in the limiting sense $X \rightarrow \infty$. Show that $q_2 = -q_1$ follows from $e_2 = -e_1$, and that the capacitance parameter per unit length is then given by

$$C = \frac{q_1}{e_1 - e_2} = \frac{\epsilon}{4 \ln \frac{d}{r}}.$$

1-3. Assume the conductors of Problem 1-2 to carry longitudinally uniform currents of i_1 and i_2 amperes in coincident reference directions. If v_1 and v_2 denote the voltage drops per unit length along the respective conductors (both in the same reference direction), show that for negligible conductor resistance these may be expressed by the relations

$$\begin{aligned} v_1 &= l_{11} i_1' + l_{12} i_2' \\ v_2 &= l_{21} i_1' + l_{22} i_2' \end{aligned}$$

in which the coefficients (called the partial inductance coefficients) are given by

$$l_{11} = l_{22} = 2\mu \ln \frac{1}{r}; \quad l_{12} = l_{21} = 2\mu \ln \frac{1}{d},$$

and the primes on the currents indicate differentiation with respect to time. In

manner similar to the logarithmic potentials in Problem 1-2, the partial flux linkages $\varphi_{rs} = l_{rs} i_s$ are determined to within a common additive constant and are referred to a common datum in the same limiting sense. Show that $i_2' = -i_1'$ follows from $v_2 = -v_1$ and for sinusoidal currents $i_2 = -i_1$. Under these conditions derive the inductance parameter per unit length and show that it is given by

$$L = 4\mu \ln \frac{d}{r}.$$

1-4. Consider a longitudinally uniform transmission line consisting of six identical conductors. The cross-sectional pattern forms a rectangular grid of which the groups of three conductors each lying in the same longitudinal plane form the down and return paths respectively for the net current which divides between the conductors in a group. The charges per unit length (all assumed positive) and the conductor currents (all in coincident reference directions) are denoted by $q_1, q_2, \dots, q_6$ and $i_1, i_2, \dots, i_6$, respectively. The common conductor radius is r , and the perpendicular distances between conductor axes are d_{rs} , ($r, s = 1, 2, \dots, 6$) where the subscripts refer to the similarly numbered conductors. Again assume $r \ll d_{rs}$. Show that if we define the set of partial elastance coefficients

$$s_{rs} = \frac{2}{\epsilon} \ln \frac{1}{d_{rs}}, \quad \text{with } d_{rr} = r,$$

and the set of partial inductance coefficients

$$l_{rs} = 2\mu \ln \frac{1}{d_{rs}}, \quad \text{with } d_{rr} = r,$$

the following systems of linear equations are obtained for the conductor potentials $e_1, \dots, e_6$ and voltage drops per unit length $v_1, \dots, v_6$ (neglecting conductor resistances)

$$\sum_{s=1}^6 s_{rs} q_s = e_r; \quad (r = 1, 2, \dots, 6), \quad (a)$$

and

$$\sum_{s=1}^6 l_{rs} i_s' = v_r; \quad (r = 1, 2, \dots, 6). \quad (b)$$

Show that the longitudinal uniformity demands that

$$\begin{aligned} e_1 = e_2 = e_3 = -e_4 = -e_5 = -e_6 = e \\ v_1 = v_2 = v_3 = -v_4 = -v_5 = -v_6 = v \end{aligned}$$

and from this total of relations find the division of the charges and currents (assumed sinusoidal) between the conductors in each group, noting that

$$\begin{aligned} q &= q_1 + q_2 + q_3 \\ i &= i_1 + i_2 + i_3 \end{aligned}$$

are the total charge and current per group but that these do *not* divide uniformly (skin and proximity effects for the conductor groups). Show that the division of charges is the same as that for the currents, and that this is due to having neglected the conductor resistances. Finally, derive expressions (using determinant notation) for the inductance and capacitance parameters based upon the relations

$$L = \frac{2v}{i'} \quad \text{and} \quad C = \frac{q}{2e};$$

ETHERFORCE

and show that in this non-dissipative case

$$LC = \varepsilon\mu.$$

1-5. For Problem 1-4 show that the consideration of the conductor resistance leads to the same problem formulation with the one difference that in the system equations (b) the terms $l_{rr}i_r'$ on the principal diagonal are replaced by $(\rho_r i_r + l_r$ where ρ_r are the conductor resistances per unit length. Show that the divergence of currents is now no longer identical with that for the charges and as a consequence $LC \neq \varepsilon\mu$.

1-6. Generalize the conductor configuration of Problems 1-4 and 1-5 by considering a total of n conductors arbitrarily spaced in the cross-sectional plane, of which the first m conductors together form the down path and the remaining $n-m$ return where m is not necessarily equal to $n/2$. First neglect conductor resistance and show that the group skin and proximity effects are the same for charges and currents (assumed sinusoidal) and that then $LC = \varepsilon\mu$. Second, consider conductor resistances and show that this is in general no longer true but holds in the limit of very large frequencies. On the basis of these results prove the correctness of the statement in the text that the skin and proximity effects for a pair of large conductors is the same for charge and current when dissipation is neglected and that this is approximately true in the dissipative case for relatively high frequencies. Derive a criterion for the term "relatively high."

1-7. Apply the results of the preceding problems to numerical cases and various simple cross-sectional conductor patterns (grid or regular polygon arrangements) assigned by the instructor or chosen by the individual. In each case determine the product LC and the quotient L/C (the latter determines the characteristic impedance of the line as discussed in the succeeding chapters).

CHAPTER II

THE STEADY-STATE SOLUTION TO THE LONG-LINE EQUATIONS

1. **The general linear character of the equations.** In this chapter we wish to discuss primarily the method of obtaining the steady-state behavior of the long line when it is excited by means of a simple harmonic driving force. This we shall do on the basis of the engineering formulation of the equilibrium conditions discussed in the previous chapter, and expressed by the equations (26), which we repeat here for convenience, omitting the subscripts "naught" since the point $x = 0$ may be considered arbitrarily located

$$\left. \begin{aligned} \frac{\partial e}{\partial x} + L \frac{\partial i}{\partial t} + Ri &= 0 \\ \frac{\partial i}{\partial x} + C \frac{\partial e}{\partial t} + Ge &= 0 \end{aligned} \right\} \quad (29)$$

There is one outstanding characteristic which these equations have in common with those for lumped-constant networks, namely, that the individual terms are all linear in e and i . This linearity of the equations is a valuable property from the standpoint of the ease of obtaining solutions when several forces are impressed either simultaneously or in succession. In fact, the validity of the superposition principle, the usefulness of which has been appreciated in the treatment of lumped-constant networks, depends upon the linearity of the system (29). It is well to bear in mind, however, that the equilibrium conditions are linear only in so far as the approximations involved in the engineering formulation are justified, because the linearity would obviously be destroyed as soon as the functional dependence of L or C upon the current or voltage were taken into account.

It may also be well to point out to the student an essential difference between the general form of the equilibrium conditions (29) and those for the usual lumped system. In the latter the currents and voltages are functions of time only, their location in the various meshes of the system being indicated by suitable subscripts. The equilibrium conditions there are given by as many simultaneous equations as there are independent meshes. Here the system is not broken up into a discrete number of meshes but is in every respect similar to a continuous medium.

The outstanding characteristic of such a medium is that the dependent variables (in our case e and i) are continuous functions of the point location within the medium instead of being discontinuous as are, for example, the currents as we go from mesh to mesh in a lumped network. In the present instance the system is continuous only with respect to one dimension, namely, x , which is the distance along the line. This is so because we have described the condition of the system wholly in terms of the voltage and current functions and have thus, formally at least, suppressed the general space-wave character of the phenomenon.

This continuous or uniformly distributed character of the system brings with it at once an inherent simplification as well as an apparent complication in the expression of its equilibrium condition. In the light of the lumped-network procedure, we would here be faced with a system with virtually an infinite number of meshes of infinitesimal size and a description of its dynamic equilibrium would therefore entail an infinite number of "mesh equations." The fact that the system is uniform, except for its boundaries (the ends of the line), makes it possible to avoid much needless writing of equations, for all those for any internal infinitesimal mesh are identical in form. It is therefore only necessary to express the equilibrium conditions for a single infinitesimal or differential internal mesh, as is done by the equations (29). This is an obvious simplification in form. The complication lies in the fact that we are mathematically faced with a novel feature in our differential equations. First of all, the variables e and i are functions of the two quantities x and t ; hence partial derivatives appear for the reason that distance and time are involved simultaneously, and in order to consider their effects separately the partial derivative scheme is resorted to. Secondly, the question as to whether the system is being driven or is force-free is not contained in these equations. They are homogeneous in either case. In the lumped system the equations were homogeneous only in the force-free case. The excited condition led to a system of inhomogeneous equations. This situation, which is a bit puzzling at first, is due to the fact that the equations (29) really do not state the equilibrium conditions of the entire system but only of the internal portion, i.e., exclusive of the line ends at which the exciting force takes hold. It is, therefore, natural that a description of the mode of excitation should not appear in these equations. Mathematically this means that the effect of the presence or absence of excitation at the line ends will subsequently have to be taken into account, i.e., after formal solutions to the system (29) have been found. This latter process is spoken of as taking the effect of *boundary conditions* into account. For the lumped system we are already familiar with a

similar process regarding initial conditions which are in a sense the temporal boundaries about which the differential equations contain no information. Our present problem, involving partial instead of total differential equations, is therefore complicated only by the fact that boundary as well as initial conditions must be subsequently met by the formal solutions, and it is only in the boundary conditions that the effect of an exciting force will be felt.

The question of meeting initial conditions arises, of course, only in connection with the determination of the transient portion of the solution and hence only when we are interested in the complete solution to a given problem. This part of the picture is the same as for the lumped-constant problem and needs no further comment here. In the present chapter we are interested only in obtaining the steady-state solution for harmonic excitation, and hence the question as to initial conditions does not enter into the discussion. In the following we shall give first a general mode of procedure which may be followed for obtaining either the forced or the force-free behavior, and then discuss a shorter process of integration particularly adapted to obtaining the harmonically excited steady state alone.

2. Formal solution by separation of variables. In attempting to find solutions to the system (29), it might seem obvious that the first step is to separate e and i , i.e., obtain separate equations for the voltage and current. Although this procedure is perfectly good, and we shall begin by adopting it, it is not necessarily the most direct method of approach as will be seen later.

The elimination of i may be accomplished as follows: Let the first equation (29) be differentiated partially with respect to x and the second with respect to t so as to get

$$\frac{\partial^2 e}{\partial x^2} + L \frac{\partial^2 i}{\partial t \partial x} + R \frac{\partial i}{\partial x} = 0, \quad (30)$$

$$\frac{\partial^2 i}{\partial x \partial t} + C \frac{\partial^2 e}{\partial t^2} + G \frac{\partial e}{\partial t} = 0. \quad (31)$$

Now since

$$\frac{\partial^2 i}{\partial x \partial t} = \frac{\partial^2 i}{\partial t \partial x},$$

we obtain from (31)

$$\frac{\partial^2 i}{\partial t \partial x} = -C \frac{\partial^2 e}{\partial t^2} - G \frac{\partial e}{\partial t}, \quad (32)$$

while from the second equation (29) we see that

$$\frac{\partial i}{\partial x} = -C \frac{\partial e}{\partial t} - Ge. \quad (33)$$

ETHERFORCE

Using these in (30) we get

$$\frac{\partial^2 e}{\partial x^2} = LC \frac{\partial^2 e}{\partial t^2} + (RC + LG) \frac{\partial e}{\partial t} + RGe, \quad (34)$$

which is the desired equation in e alone.

In a similar manner, by differentiating the first equation (29) partially with respect to t and the second with respect to x , we obtain an identical equation for the current, namely

$$\frac{\partial^2 i}{\partial x^2} = LC \frac{\partial^2 i}{\partial t^2} + (RC + LG) \frac{\partial i}{\partial t} + RGi. \quad (35)$$

The equations (34) and (35) are called *wave equations* for the reason that the solutions for e and i have the character of traveling waves, as will be seen later.

The next step in the process of solution involves the separation of the independent variables x and t which are simultaneously involved. This separation, in wave equations of this type, can always be done by assuming a form of solution which is quite logical on the basis of our experience with lumped networks. There we found that all the current and voltages are equal to the same function of time with the exception that different amplitudes and phases are involved in the different meshes. Therefore, it is logical to assume here that the voltage, for example, is given by a time function whose amplitude is a function of x only. The same argument holds with respect to the current. This is the same as to say that the voltage or current ought to be given by the product of a function of distance and a function of time.

In order to illustrate the generality of the method, let us assume for the voltage equation (34) that

$$e(x, t) = g(x) \cdot h(t). \quad (36)$$

Indicating differentiation by means of primes, substitution into (34) gives

$$g''h = LCgh'' + (RC + LG)gh' + RGgh. \quad (37)$$

Dividing through by gh this becomes

$$\frac{g''}{g} = LC \frac{h''}{h} + (RC + LG) \frac{h'}{h} + RG. \quad (38)$$

Here we see that the left-hand side is a function of x only while the right-hand side is a function of t only. But a function of x can equal

a function of t for all values of x and t only when each reduces to the same constant. Denoting this constant by α^2 , we therefore have

$$g'' - \alpha^2 g = 0, \quad (39)$$

and

$$h'' + \left(\frac{R}{L} + \frac{G}{C}\right)h' + \left(\frac{RG - \alpha^2}{LC}\right)h = 0. \quad (40)$$

These are now total differential equations in x and t respectively. They are linear equations with constant coefficients of the general type with which we are already familiar from our experience with lumped networks. Hence we know without further comment that they are satisfied by exponential functions, as may be verified through substitution. This is nothing surprising so far as the time function is concerned. Indeed, we should have suspected this from the beginning, but we wished here merely to show that the above process of separating the variables x and t is perfectly general and has nothing to do with the fact that the solutions are of the exponential form. That is why we chose to denote the distance and time functions by means of the noncommittal letters g and h , respectively.

Knowing that the time function is an exponential, we can arrive at a result more quickly by making this assumption directly in (34) and (35). To show how this proceeds let us assume

$$\left. \begin{aligned} e(x, t) &= E(x) \cdot e^{pt} \\ i(x, t) &= I(x) \cdot e^{pt} \end{aligned} \right\} \quad (41)$$

These are really the same kind of assumptions as were made for lumped networks, except that here the amplitudes E and I are not constants but continuous functions of the distance along the line. Our heuristic method of procedure would therefore have suggested these forms from the beginning. Substitution into the wave equations (34) and (35) gives, after rearrangement of terms and cancelation of the time function e^{pt} ,

$$\left. \begin{aligned} \frac{d^2 E}{dx^2} - \alpha^2 E &= 0 \\ \frac{d^2 I}{dx^2} - \alpha^2 I &= 0 \end{aligned} \right\} \quad (42)$$

where we have put

$$\alpha^2 = (R + Lp)(G + Cp). \quad (43)$$

Hence the assumptions (41) will be valid and useful provided the distance functions E and I can be determined so as to satisfy the differ-

ential equations (42) as well as any special conditions that we wish to impose at the boundaries.

The equations (42) are obviously satisfied by exponentials. For example, let us assume

$$E(x) = Ae^{mx}. \quad (44)$$

Then

$$\frac{d^2 E}{dx^2} = m^2 Ae^{mx} = m^2 E,$$

so that the first equation (42) becomes

$$(m^2 - \alpha^2)E = 0,$$

which demands that for a non-vanishing E we must have

$$m = \pm \alpha.$$

Hence the complete expression for E is given by the sum of two such terms as (44), one for each value of m , namely,

$$E(x) = A_1 e^{-\alpha x} + A_2 e^{\alpha x}. \quad (45)$$

Similarly we get

$$I(x) = B_1 e^{-\alpha x} + B_2 e^{\alpha x}. \quad (46)$$

These are the formal expressions for the distance functions.

We are now ready to take into account the fact that the system is being driven by a harmonic force at one end of the line. Let us assume here a generator having the following terminal voltage function

$$e_g = \Re [E_g e^{j\omega t}], \quad (47)$$

where ω is the angular frequency, and the $\Re$ sign has the usual significance of taking the real part of whatever it operates upon, just as in the treatment of lumped networks. This sign may, of course, be dropped temporarily and later inserted before interpreting the instantaneous voltage and current functions.

We now make use of our experience with linear systems by saying that, when such a system is driven by a harmonic force at any point then the steady-state voltage and current response at all points will be harmonic and of the same frequency as the driving force. This argument makes it clear that in the expressions (41), we must put $p = j\omega$ and thus write

$$\left. \begin{aligned} e(x, t) &= \Re [E(x) \cdot e^{j\omega t}] \\ i(x, t) &= \Re [I(x) \cdot e^{j\omega t}] \end{aligned} \right\} \quad (48)$$

These expressions, together with (45) and (46), constitute the complete formal solutions to the harmonically excited steady state.

3. Direct solution for the steady state. The above procedure for obtaining the steady-state solutions is perhaps more lengthy and explicit than necessary, but we have given it in the hope that it might afford the reader a more thorough insight into the mechanism involved. Then too, owing to its more general mode of approach, it will yield the transient or force-free behavior as well as the steady-state which is our more immediate objective. In a later chapter we shall discuss the transient solution in detail, and will find that the forms (41) hold there also except that the values for p , instead of being equal to the value $j\omega$, are the natural modes of the system—again just as in lumped networks.

However, if we are interested only in a harmonic steady state, and have given a driving force of the form (47), then our experience should at once tell us that the forms (48) are reasonable, and that we ought to substitute these directly into the given equations (29) without bothering to separate e and i . Dropping the $\Re$ sign, these substitutions immediately yield

$$\left. \begin{aligned} \frac{dE}{dx} + (R + jL\omega)I &= 0 \\ \frac{dI}{dx} + (G + jC\omega)E &= 0 \end{aligned} \right\}, \quad (49)$$

where the factor $e^{j\omega t}$ has been canceled. Here E or I may easily be eliminated by differentiating either equation with respect to x and substituting for the single derivative term from the other equation. The reader will thus see that we obtain

$$\left. \begin{aligned} \frac{d^2 E}{dx^2} - \alpha^2 E &= 0 \\ \frac{d^2 I}{dx^2} - \alpha^2 I &= 0 \end{aligned} \right\}, \quad (42a)$$

with

$$\alpha^2 = (R + jL\omega)(G + jC\omega), \quad (43a)$$

which is the same as (43) with $p = j\omega$. The pair (42a) is identical with (42), and hence the solutions (45) and (46) follow as before.

4. Boundary conditions and the evaluation of integration constants. The next step is to adapt the formal solutions (45) and (46) to the boundary conditions of our problem. For this purpose we have available the integration constants A_1 , A_2 , B_1 , and B_2 . Here we immediately notice an apparent defect, however. Our line has obviously only two ends, at each of which one relationship between voltage, current, and terminal impedance may be set up. This process will therefore suffice to determine only two constants, whereas apparently we have four. Obviously only two of these can be arbitrary. This we recognize to

ETHERFORCE

be so if we realize that the relations (49) definitely link the voltage and current amplitudes. This fact we have not yet made use of but shall proceed to do so immediately.

Substituting (45) and (46) into (49), and writing for brevity

$$\left. \begin{aligned} y &= G + jC\omega \\ z &= R + jL\omega \end{aligned} \right\}, \quad (50)$$

we find after grouping terms with $e^{-\alpha x}$ and $e^{\alpha x}$

$$\left. \begin{aligned} (zB_1 - \alpha A_1)e^{-\alpha x} + (zB_2 + \alpha A_2)e^{\alpha x} &= 0 \\ (yA_1 - \alpha B_1)e^{-\alpha x} + (yA_2 + \alpha B_2)e^{\alpha x} &= 0 \end{aligned} \right\}, \quad (51)$$

or noting (43a) and (50) we have

$$\left. \begin{aligned} (Z_0B_1 - A_1)e^{-\alpha x} + (Z_0B_2 + A_2)e^{\alpha x} &= 0 \\ (A_1 - Z_0B_1)e^{-\alpha x} + (A_2 + Z_0B_2)e^{\alpha x} &= 0 \end{aligned} \right\}, \quad (52)$$

where the symbol

$$Z_0 = \sqrt{\frac{z}{y}} = \sqrt{\frac{R + jL\omega}{G + jC\omega}} \quad (53)$$

has been introduced. Since these relations are to hold for all value of x , the coefficients of $e^{-\alpha x}$ and $e^{\alpha x}$ must vanish separately. Making use of this fact we see that the equations (52) are consistent in yielding

$$\left. \begin{aligned} B_1 &= \frac{A_1}{Z_0} \\ B_2 &= \frac{A_2}{-Z_0} \end{aligned} \right\}, \quad (54)$$

so that we now have

$$\left. \begin{aligned} E(x) &= A_1e^{-\alpha x} + A_2e^{\alpha x} \\ I(x) &= \frac{A_1}{Z_0}e^{-\alpha x} - \frac{A_2}{Z_0}e^{\alpha x} \end{aligned} \right\} \quad (55)$$

for the voltage and current amplitudes. These contain only two arbitrary constants A_1 and A_2 , which may now be evaluated from boundary conditions in the following manner.

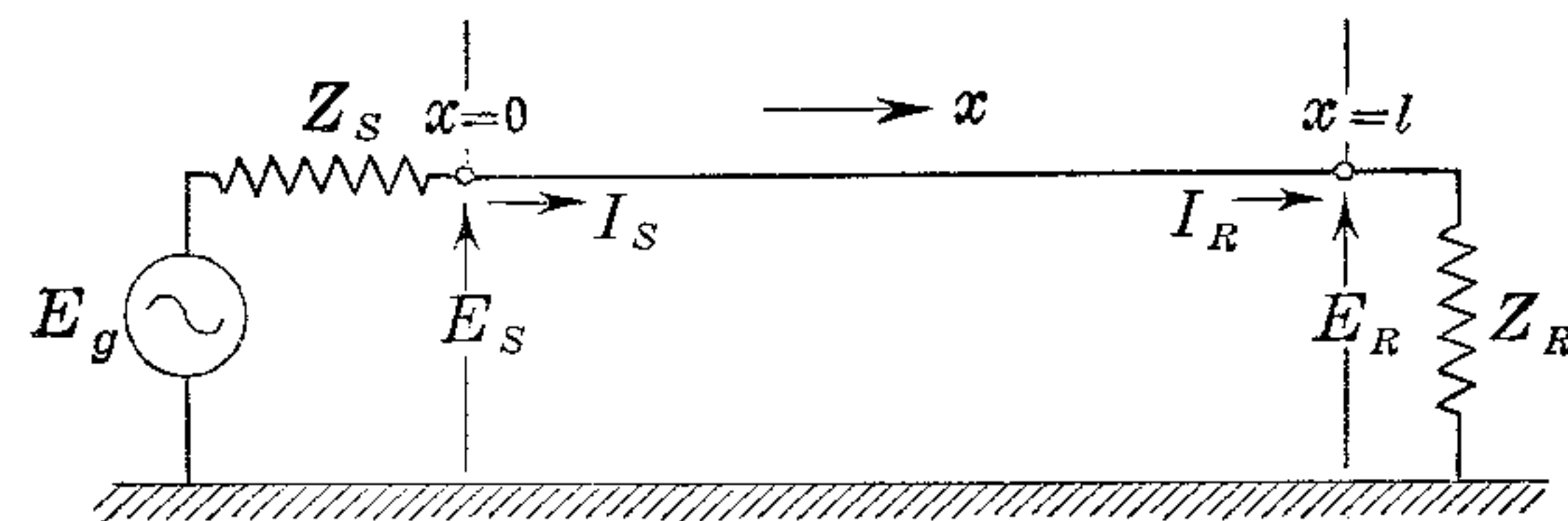


FIG. 4.—Single-line diagram of the transmission line showing terminal conditions and assumed reference directions.

of excitation. For this figure the so-called single-line diagram is used which replaces the return conductor by a ground plane. This is merely

for the purpose of simplifying the appearance of the figure. The sending end of the line, where the generator is located, is denoted by $x = 0$; the receiving end is denoted by $x = l$; i.e., distance is measured positively from sending to receiving end, and the line length is denoted by l . Z_S and Z_R are the sending- and receiving-end impedances, respectively, for the angular frequency ω . The positive assumed directions for current are from left to right, and the positive directions for voltage are from the ground plane to the line.

On this basis we may write the following equations at the sending and receiving ends, respectively. The subscripts S and R refer to these ends.

$$\left. \begin{aligned} E_S + I_S Z_S &= E_g \\ E_R - I_R Z_R &= 0 \end{aligned} \right\}. \quad (56)$$

By (55) we see that

$$\left. \begin{aligned} E_S &= A_1 + A_2 \\ E_R &= A_1e^{-\alpha l} + A_2e^{\alpha l} \\ I_S &= \frac{1}{Z_0}(A_1 - A_2) \\ I_R &= \frac{1}{Z_0}(A_1e^{-\alpha l} - A_2e^{\alpha l}) \end{aligned} \right\}. \quad (57)$$

Substituting these into (56), the boundary conditions take the form

$$\left. \begin{aligned} (Z_0 + Z_S)A_1 + (Z_0 - Z_S)A_2 &= Z_0E_g \\ (Z_0 - Z_R)e^{-\alpha l}A_1 + (Z_0 + Z_R)e^{\alpha l}A_2 &= 0 \end{aligned} \right\}, \quad (56a)$$

by means of which A_1 and A_2 may be evaluated.

Denoting the determinant of this system of simultaneous equations by D , we have

$$D = (Z_0 + Z_S)(Z_0 + Z_R)e^{\alpha l} - (Z_0 - Z_S)(Z_0 - Z_R)e^{-\alpha l}, \quad (58)$$

and hence

$$\left. \begin{aligned} A_1 &= \frac{(Z_R + Z_0)Z_0e^{\alpha l}E_g}{D} \\ A_2 &= \frac{(Z_R - Z_0)Z_0e^{-\alpha l}E_g}{D} \end{aligned} \right\}. \quad (59)$$

Substituting these back into (55), the voltage and current amplitudes become

$$E(x) = \frac{Z_0E_g[(Z_R + Z_0)e^{\alpha(l-x)} + (Z_R - Z_0)e^{-\alpha(l-x)}]}{(Z_S + Z_0)(Z_R + Z_0)e^{\alpha l} - (Z_S - Z_0)(Z_R - Z_0)e^{-\alpha l}}, \quad (60)$$

and

$$I(x) = \frac{E_g[(Z_R + Z_0)e^{\alpha(l-x)} - (Z_R - Z_0)e^{-\alpha(l-x)}]}{(Z_S + Z_0)(Z_R + Z_0)e^{\alpha l} - (Z_S - Z_0)(Z_R - Z_0)e^{-\alpha l}}, \quad (61)$$

which together with (48), (43a), and (53) constitute the complete steady state solution desired.

For purposes of interpretation it is convenient to modify the form of the results (60) and (61) in the following way. Introducing the symbols

$$r_S = \frac{Z_S - Z_0}{Z_S + Z_0} \quad (6)$$

and

$$r_R = \frac{Z_R - Z_0}{Z_R + Z_0} \quad (6)$$

the expressions (60) and (61) may be put in the form

$$E(x) = \frac{Z_0 E_g (e^{\alpha(l-x)} + r_R e^{-\alpha(l-x)})}{(Z_S + Z_0)(e^{\alpha l} - r_S r_R e^{-\alpha l})} \quad (60)$$

and

$$I(x) = \frac{E_g (e^{\alpha(l-x)} - r_R e^{-\alpha(l-x)})}{(Z_S + Z_0)(e^{\alpha l} - r_S r_R e^{-\alpha l})} \quad (61)$$

The quantities r_S and r_R may be given a useful physical interpretation. This we wish to take up in more detail now.

5. Wave character of the solution. In this section we wish to show that the steady-state solution found above may be interpreted as physically representing a pair of waves traveling in opposite directions with the same constant velocity. In order to do this it is convenient to make use of the following further notation. Let

$$E^+ = \frac{Z_0 E_g}{(Z_S + Z_0)(e^{\alpha l} - r_S r_R e^{-\alpha l})} \quad (62)$$

and

$$E^- = r_R E^+, \quad (63)$$

also

$$I^+ = \frac{E^+}{Z_0} \quad (64)$$

and

$$I^- = \frac{E^-}{-Z_0} = -r_R I^+. \quad (65)$$

Then (60a) and (61a) become

$$\left. \begin{aligned} E(x) &= E^+ e^{\alpha(l-x)} + E^- e^{-\alpha(l-x)} \\ I(x) &= I^+ e^{\alpha(l-x)} + I^- e^{-\alpha(l-x)} \end{aligned} \right\} \quad (66)$$

Now it is necessary to recognize that the quantity α , defined by (43a), is complex, and hence may be written

$$\alpha = \alpha_1 + j\alpha_2. \quad (43b)$$

α is called the **propagation function** of the line, because, as we shall see, it governs the propagation properties of the system. The real part of α , namely α_1 , governs the attenuation of the waves, and hence is called the **attenuation function**, while α_2 will be seen to determine the variation of phases with distance and hence is given the name **phase function**.

One more set of definitions is necessary before the wave expressions may be set up. This involves the complex character of the quantities (64), (64a), (65), and (65a). Here we shall write

$$\left. \begin{aligned} E^+ &= |E^+| \underline{\psi^+} \\ E^- &= |E^-| \underline{\psi^-} \\ I^+ &= |I^+| \underline{\varphi^+} \\ I^- &= |I^-| \underline{\varphi^-} \end{aligned} \right\} \quad (67)$$

With these notations, the amplitudes (66) may now be substituted into (48) and the real parts evaluated. We find

$$e(x,t) = |E^+| e^{\alpha_1(l-x)} \cdot \cos [\alpha_2(l-x) + (\omega t + \psi^+)] \\ + |E^-| e^{-\alpha_1(l-x)} \cdot \cos [\alpha_2(l-x) - (\omega t + \psi^-)], \quad (68)$$

and

$$i(x,t) = |I^+| e^{\alpha_1(l-x)} \cdot \cos [\alpha_2(l-x) + (\omega t + \varphi^+)] \\ + |I^-| e^{-\alpha_1(l-x)} \cdot \cos [\alpha_2(l-x) - (\omega t + \varphi^-)]. \quad (69)$$

The first terms in these expressions represent cosine curves travelling in the positive direction with amplitudes which decay exponentially in the direction of motion. These terms are called the *incident waves*. The second terms are cosine curves traveling in the opposite direction with the same velocity and decaying exponentially in that direction. They are called the *reflected waves*, it being supposed that their presence is due to the fact that a portion of the incident wave is reflected at the receiving end. Noting (64a) and (65a), we see that the magnitudes of the reflected waves at the receiving end (point of reflection) equal the magnitudes of the incident waves at that point multiplied by the magnitude of the factor r_R which is, therefore, called the **reflection coefficient** of the receiving end. The factor r_S , which is called the reflection coefficient of the sending end, does not play the same rôle at that end because the source is assumed located there. If we had considered generators at either end, we would have found a similar reflection phenomenon taking place at the sending end also. The student may set up the boundary conditions for this condition and evaluate the result as an exercise.

The reflection coefficients given by (62) and (63) evidently depend upon the discrepancy between the respective terminal impedance and

the quantity Z_0 defined by (53), which, incidentally, also has the dimensions of an impedance. The latter is characteristic of the line itself and hence is given the name **characteristic impedance**. A physical interpretation of this quantity alone will be given later when a full appreciation of the forms (62) and (63) for the reflection coefficient will be attained.

In order to see more clearly that the terms in (68) and (69) referred to above really are traveling cosine curves, the reader may first consider the following simpler expressions:

$$\text{and } \left. \begin{aligned} \cos(\alpha_2 x - \omega t) \\ \cos(\alpha_2 x + \omega t) \end{aligned} \right\} \quad (70)$$

which represent the essential features of the incident and reflected waves respectively, the phases $\alpha_2 l$ and ψ^+ , etc., having nothing to do with the general characteristics of these terms. The student should plot the terms (70) as functions of x for successive increasing values of t , obtaining sets of curves for each term. Thus it will become clear that the first progresses in the positive, and the second in the negative direction. He will then appreciate more fully the wave character of expressions (68) and (69). He should also convince himself of the fact that each component wave (incident and reflected) damps exponentially in the direction of its travel, and that the point of reflection is at the receiving end. There the amplitude of the reflected wave is fixed by that of the incident wave and the reflection coefficient. Unless this point is clearly appreciated, it might be thought that, since the reflected wave grows exponentially in the positive x -direction, its amplitude may increase beyond all bounds.

The only reason for putting the steady-state solution into this form is to aid in an attempt to visualize the resulting phenomenon. Whether or not waves are really present, or what the waves consist of, is altogether immaterial. The wave interpretation should be regarded as a mathematical picture rather than as a physical fact. After all, we can say with certainty only that the results, so far as measurable voltage and current are concerned, *are as though* waves, such as we have described, are present on the system. For this we have mathematical proof, namely in the results (68) and (69). The harmonic steady state is therefore, as though it were due to the continuous passage of waves in opposite directions at all points in the line, the net voltage or current at any point being given by the linear superposition of the separate contributions of the incident and reflected wave components respectively, where proper attention is paid to the time and space phases of

these components. Thus, in the complex voltage and current amplitudes as expressed by (66), the first terms represent the complex amplitudes of the incident waves at any point x , and the second terms represent the complex amplitudes of the reflected waves. The *vector sum* of the complex amplitudes of the incident and reflected waves at any point gives the net complex amplitude E or I at that point.

These relations (66) also make it clear that the point of reflection is at the receiving end, i.e., that the component amplitudes E^+ and E^- as well as I^+ and I^- refer to that point. To see this, let $x = l$ (which fixes our attention upon the receiving end). Then (66) becomes

$$\left. \begin{aligned} E(l) &= E_R = E^+ + E^- \\ I(l) &= I_R = I^+ + I^- \end{aligned} \right\} \quad (66a)$$

At the sending end, on the other hand, $x = 0$, and we have

$$\left. \begin{aligned} E(0) &= E_S = E^+ e^{\alpha l} + E^- e^{-\alpha l} \\ I(0) &= I_S = I^+ e^{\alpha l} + I^- e^{-\alpha l} \end{aligned} \right\} \quad (66b)$$

which makes it clear that the complex amplitudes of the incident waves at the sending end are $E^+ e^{\alpha l}$ and $I^+ e^{\alpha l}$, while those for the reflected waves at that end are $E^- e^{-\alpha l}$ and $I^- e^{-\alpha l}$. This also makes it clear that the ratio of an incident wave amplitude at the sending end to that wave amplitude at the receiving end is $e^{\alpha l}$, and that the corresponding ratio for the reflected wave amplitudes is $e^{-\alpha l}$. The factor $e^{\alpha l}$ or $e^{-\alpha l}$, therefore, represents the change which a component wave amplitude undergoes (both as to magnitude and phase on account of the complex character of α) as it traverses the length of the line. The propagation properties of the line, as they affect the individual wave components, are bound up—so to speak—in the factor $e^{\alpha l}$, or, we may say, in α , which justly, therefore, deserves the name of propagation function as already mentioned.

Another important point which the student should not fail to appreciate in this connection is brought out by the relations (65) and (65a), which show that the ratio of complex voltage and current amplitudes equals $+Z_0$ for the incident waves, and $-Z_0$ for the reflected waves; i.e., the *component* voltage and current wave amplitudes are simply related by the characteristic impedance of the line in the Ohm's law sense, with the novelty that a negative relationship appears with the reflected waves. This simple relationship between voltage and current, however, exists only for the component waves, and the student should not make the mistake of imagining that it, therefore, holds for the net voltage and current at any point! That will be true only under special conditions which cause the reflected wave to be absent. Then the net values equal the incident values so that the net voltage

and current at any point are simply related by the characteristic impedance. This important condition we wish now to discuss in more detail.

6. Characteristic impedance. We mentioned earlier that we would give the reader a more concrete physical notion of how the characteristic impedance (53) may be interpreted. To do this let us consider the expressions (60a) and (61a) for the complex voltage and current amplitudes. In these expressions let us fix our attention particularly upon those terms in which the factor $e^{-\alpha l}$ appears. Recognizing the complex character of α , this factor becomes

$$e^{-\alpha_1 l} \cdot e^{-j\alpha_2 l}.$$

Now suppose that the attenuation function α_1 has some finite value and that the length of the line becomes very large. The factor $e^{-\alpha}$ then becomes very small. If the line were infinitely long, this factor would vanish altogether, and hence the terms with which it is associated in (60a) and (61a) would also vanish. These are the terms in which r_R appears. When this happens, the resulting expressions for $E(x)$ and $I(x)$ become independent of the line length l because the factors $e^{\alpha l}$ cancel out. This cancelation is perfectly admissible even though l is tending toward infinity. Hence we can say

$$\lim_{l \rightarrow \infty} [E(x)] = \frac{Z_0 E_g}{Z_S + Z_0} \cdot e^{-\alpha x}, \quad (71)$$

and

$$\lim_{l \rightarrow \infty} [I(x)] = \frac{E_g}{Z_S + Z_0} \cdot e^{-\alpha x}. \quad (72)$$

Another way of showing this is to factor $e^{\alpha l}$ out of the numerator and denominator of the expressions (60a) and (61a) so that these become

$$E(x) = \frac{Z_0 E_g e^{-\alpha x}}{Z_S + Z_0} \left\{ \frac{1 + r_R e^{-2\alpha(l-x)}}{1 - r_S r_R e^{-2\alpha l}} \right\}, \quad (60b)$$

$$I(x) = \frac{E_g e^{-\alpha x}}{Z_S + Z_0} \left\{ \frac{1 - r_R e^{-2\alpha(l-x)}}{1 - r_S r_R e^{-2\alpha l}} \right\}. \quad (61b)$$

For large values of l the terms involving $e^{-2\alpha l}$ become small, so that we may expand the bracket terms and neglect all but the linear terms in $e^{-2\alpha l}$. This gives

$$E(x) = \frac{Z_0 E_g e^{-\alpha x}}{Z_S + Z_0} \left\{ 1 + r_R(r_S + e^{2\alpha x})e^{-2\alpha l} + \dots \right\}, \quad (60c)$$

$$I(x) = \frac{E_g e^{-\alpha x}}{Z_S + Z_0} \left\{ 1 + r_R(r_S - e^{2\alpha x})e^{-2\alpha l} + \dots \right\}. \quad (61c)$$

In the limit $l \rightarrow \infty$ all terms in the bracketed series vanish for finite

values of x except the first which are unity, and thus the forms (71) and (72) are obtained. Using these in (48) we see that for an infinitely long line

$$\left. \begin{aligned} e(x,t) &= \operatorname{Re} \left[\frac{Z_0 E_g}{Z_S + Z_0} \cdot e^{-\alpha x + j\omega t} \right] \\ i(x,t) &= \operatorname{Re} \left[\frac{E_g}{Z_S + Z_0} \cdot e^{-\alpha x + j\omega t} \right] \end{aligned} \right\}, \quad (73)$$

or if we let

$$\frac{Z_0 E_g}{Z_S + Z_0} = \left| \frac{Z_0 E_g}{Z_S + Z_0} \right| \underline{\psi'}, \quad (74)$$

and

$$\frac{E_g}{Z_S + Z_0} = \left| \frac{E_g}{Z_S + Z_0} \right| \underline{\phi'}, \quad (75)$$

then

$$\left. \begin{aligned} e(x,t) &= \left| \frac{Z_0 E_g}{Z_S + Z_0} \right| \cdot e^{-\alpha_1 x} \cdot \cos [\alpha_2 x - \omega t - \psi'] \\ i(x,t) &= \left| \frac{E_g}{Z_S + Z_0} \right| \cdot e^{-\alpha_1 x} \cdot \cos [\alpha_2 x - \omega t - \phi'] \end{aligned} \right\}. \quad (73a)$$

These represent waves traveling in the positive direction. The net voltage and current are, therefore, represented by their incident wave components alone. The reflected waves are absent, as might have been predicted since the point of reflection is infinitely far away. In complete agreement with this situation, we see from (71) and (72) that for this case

$$\frac{E(x)}{I(x)} = Z_0, \quad (76)$$

i.e., the impedance looking into the infinitely long line at any point is equal to Z_0 ! This gives us a very convenient physical interpretation for the characteristic impedance.

With this in mind, the significance of the reflection coefficients (62) and (63) becomes somewhat clearer also. For example, it is obviously immaterial whether the line is infinitely long or whether it is finite and terminated in a lumped impedance Z_R which is equal to the characteristic impedance, because we have just shown that the latter is equivalent to an infinitely long section of line. Consistent with this idea, we see by (62) that letting $Z_R = Z_0$ makes $r_R = 0$, i.e., there is no reflection at the receiving end under these conditions. If we refer to the expressions (60a) and (61a) for the net voltage and current amplitudes, we see that setting $r_R = 0$ reduces these to the forms (71)

and (72) which were obtained for the infinitely long line. Hence we again see that whether we consider the line infinitely long, or finite but terminated in Z_0 , the results are the same. The voltage and current expressions reduce to a single exponential term in either case, and are simply related by the characteristic impedance. In practice this condition is highly desirable for reasons which will be discussed in detail later.

The forms (60c) and (61c) for the voltage and current amplitudes involving series expansions in terms of the factor $e^{-2\alpha l}$ are interesting from another standpoint. If the line is very long or the attenuation high, the series will converge very rapidly for relatively small values of x . In such cases, therefore, the reflected wave may be neglected to within an approximate error of

$$|r_R(r_S \pm e^{2\alpha x})e^{-2\alpha l}| \cdot 100\%. \quad (77)$$

This is smallest at the sending end, namely

$$|r_R(r_S \pm 1)e^{-2\alpha l}| \cdot 100\%. \quad (77a)$$

For the impedance looking into the line at the sending end we have

$$\frac{E(0)}{I(0)} = Z_0(1 + 2r_R e^{-2\alpha l} + \dots), \quad (76a)$$

i.e., the impedance looking into the sending end is equal to the characteristic impedance to within an approximate error of

$$2r_R e^{-2\alpha l}.$$

This means physically that, under conditions where these approximations are justified, the sending-end impedance may be assumed equal to Z_0 regardless of how the line is terminated! This is useful in many practical cases.

7. Wave length and phase velocity. The wave interpretation, already discussed to some extent above, brings with it certain additional concepts which it is desirable to point out at this time. For this purpose we shall return to the wave solutions (68) and (69). We have mentioned that these are traveling cosine curves which damp exponentially as they go. A wave of this kind, whether it decays or not, has two distinguishing characteristics in addition to its amplitude and phase relative to some arbitrary reference. One of these is made evident if we imagine the time held constant, i.e., consider the cosine function plotted versus x for a constant value of t . If we neglect the damping for the moment, this is a periodic function, of course. It is periodic in x , i.e., distance. A certain length corresponds to a period of the cosine function. This distance obviously is that length for which $\alpha_2 x$

(the variable part of the argument) increases by 2π radians. If we denote this length by λ , we have

$$\begin{aligned} \alpha_2 \lambda &= 2\pi \\ \lambda &= \frac{2\pi}{\alpha_2}. \end{aligned} \quad (78)$$

λ is called the **wave length** or also the **spacial period** because it corresponds to a space-period of the cosine function. This result will become better oriented in the reader's mind if he will associate it with another quantity with which he has an acquaintance of longer standing, namely, the temporal period of the usual harmonic function. This is usually written

$$\tau = \frac{1}{f} = \frac{2\pi}{\omega}, \quad (79)$$

where f is the frequency in cycles per second, and ω the angular frequency. Thus λ is for space what τ is for time. Furthermore, the reader should recognize the parallel rôles played by ω and α_2 . The first has the dimensions of *radians per second*, while for the second the dimensions are *radians per mile*—or whatever unit is used for distance. We see that α_2 is for distance what ω is for time! This considerably clarifies the significance of α_2 —the phase function. It is equal to the rate at which the phase of a component wave changes with distance at any instant of time.

The analogous quantities ω and α_2 may be still further correlated in the following way. Consider, for example, the voltage wave solution. We mentioned that this is a pair of traveling waves. Suppose we now ask: How fast do these waves travel? This is the same as to say: What is the velocity of a given point on either one of the cosine curves? Fixing our attention upon a given point on one of the cosine curves as it moves is equivalent to observing what happens when the argument of that cosine as a whole remains constant while the variables x and t in it change. This may be expressed mathematically by setting the differential of the argument equal to zero. Doing this for the incident wave, we have

$$d[\alpha_2(l - x) + (\omega t + \psi^+)] = 0,$$

which gives

$$\begin{aligned} -\alpha_2 dx + \omega dt &= 0, \\ \alpha_2 dx &= \omega dt, \end{aligned}$$

or

and thus we have

$$\frac{dx}{dt} = \frac{\omega}{\alpha_2}. \quad (80)$$

ETHERFORCE

This states that, as we fix our attention upon any point on the incident wave, we observe that the rate at which distance changes with time is given by ω/α_2 . But this is the velocity with which the entire wave moves, namely,

$$v = \frac{\omega}{\alpha_2} \quad (80a)$$

for the incident wave.

For the reflected wave we form

$$d[\alpha_2(l-x) - (\omega t + \psi^-)] = 0,$$

and get

$$\frac{dx}{dt} = -\frac{\omega}{\alpha_2}, \quad (81)$$

which shows that the reflected wave moves in the negative direction with the same velocity. The velocity with which the component waves move is, therefore, the correlation factor between ω and α_2 . We can also see that this result checks dimensionally because *radians per second* divided by *radians per mile* gives *miles per second*.

The reader should bear in mind in this connection that this velocity of propagation of the component waves in the harmonic steady state is not necessarily the velocity at which electromagnetic energy is propagated along the line. We mentioned earlier that the wave concept introduced in this discussion is merely a mathematical fiction which is injected into the picture for convenience in interpreting the results. The student should not make the mistake of becoming so wedded to this mode of interpretation as to forget entirely its fictitious character. If he will stop to reflect for a moment he will realize that questions regarding the propagation of energy, or voltage and current for that matter, have no utility in connection with the steady-state discussion. Velocity of propagation in the usual sense has to do with something that has a beginning and an end, and a steady state has, by definition, neither beginning nor end. Such things as velocities of wave fronts or of wave groups are part of the transient behavior to which we shall come later. In order to avoid confusion on this point, the velocity (80a), which refers only to the fictitious steady-state waves, is referred to as the **phase velocity** since it is the velocity at which a point of constant phase, for either component wave, is propagated.

8. Hyperbolic form. In the discussion so far, the solution was, for the most part, left in the exponential form. This form will usually be found most convenient for numerical calculation. There is another form in terms of hyperbolic functions, however, which finds favor in a large part of the literature on this subject, and it is for the purpose of

acquainting the reader with its relation to the exponential form that the following is given.

The hyperbolic form follows almost immediately from the relations (60) and (61) which were obtained directly after the evaluation of integration constants. If we multiply out the factored terms in the numerator and denominator of each expression, and then factor so as to form the groups

$$\begin{aligned} e^{\alpha(l-x)} + e^{-\alpha(l-x)} &= 2 \cosh \alpha(l-x) \\ e^{\alpha(l-x)} - e^{-\alpha(l-x)} &= 2 \sinh \alpha(l-x) \\ e^{\alpha l} + e^{-\alpha l} &= 2 \cosh \alpha l \\ e^{\alpha l} - e^{-\alpha l} &= 2 \sinh \alpha l \end{aligned}$$

we obtain

$$E(x) = \frac{Z_0 E_g \{ Z_R \cosh \alpha(l-x) + Z_0 \sinh \alpha(l-x) \}}{Z_0 (Z_S + Z_R) \cosh \alpha l + (Z_S Z_R + Z_0^2) \sinh \alpha l} \quad (60d)$$

$$I(x) = \frac{E_g \{ Z_0 \cosh \alpha(l-x) + Z_R \sinh \alpha(l-x) \}}{Z_0 (Z_S + Z_R) \cosh \alpha l + (Z_S Z_R + Z_0^2) \sinh \alpha l} \quad (61d)$$

which are the desired forms.

When charts for hyperbolic functions of complex arguments are available, these results may be quite convenient for numerical calculation of the complex voltage and current amplitudes. Under ordinary circumstances, however, the exponential forms are of greater utility.

9. Symmetrical forms. In general an almost endless number of variations may be given to the form of the steady-state solution, as the reader will find if he pursues the literature on this subject somewhat. A certain amount of personal taste enters into the situation, and no two authors will be found to agree exactly as to form or notation. Obviously nothing much is to be gained by giving additional variations, except to point out one which adds much to the compactness of the resulting expressions, and may save time in numerical calculations.

The reader will probably have noticed one obvious defect in the forms given so far, and that is that they lack symmetry. The complex forms (60a) and (61a), for example, are marred by the appearance of the factors r_R and $r_S r_R$ in connection with one of the exponentials in each pair of terms of the numerator and denominator. But for this lack of symmetry, these pairs could be contracted into a single hyperbolic function. This lack of symmetry may be removed by the following simple expedient.

If we factor $r_R^{1/2}$ out of the numerator and $r_S^{1/2} r_R^{1/2}$ out of the denominator, we will have for the voltage amplitude

$$E(x) = \frac{Z_0 E_g (r_R^{-1/2} e^{\alpha(l-x)} + r_R^{1/2} e^{-\alpha(l-x)})}{r_S^{1/2} (Z_S + Z_0) (r_S^{-1/2} r_R^{-1/2} e^{\alpha l} - r_S^{1/2} r_R^{1/2} e^{-\alpha l})}, \quad (82)$$

ETHERFORCE

for which the numerator and denominator now contain symmetrical groups. In order to make this still more evident, we define the quantities ρ_S and ρ_R such that

$$\left. \begin{aligned} r_S^{1/2} &= e^{\rho_S}; \rho_S = \frac{1}{2} \ln r_S \\ r_R^{1/2} &= e^{\rho_R}; \rho_R = \frac{1}{2} \ln r_R \end{aligned} \right\}. \quad (83)$$

Then the numerator and denominator consist of symmetrical exponential groups for which we can substitute the hyperbolic functions and obtain

$$E(x) = \frac{Z_0 E_g \cosh [\alpha(l - x) - \rho_R]}{\sqrt{Z_S^2 - Z_0^2} \sinh [\alpha l - \rho_S - \rho_R]}. \quad (60e)$$

Similarly the current amplitude becomes

$$I(x) = \frac{E_g \sinh [\alpha(l - x) - \rho_R]}{\sqrt{Z_S^2 - Z_0^2} \sinh [\alpha l - \rho_S - \rho_R]}. \quad (61e)$$

In both of these the expression (62) for r_S has also been made use of. Thus the effect of the terminal impedances is taken care of by the quantities ρ_S and ρ_R in the arguments of the hyperbolic functions. These quantities are sometimes referred to as the *equivalent line angles*, it being imagined that the terminal impedances have been replaced by fictitious extensions of the line. The resulting expressions are extremely compact. If hyperbolic charts are available, these forms may save time in computation, although the equivalent line angles must first be determined from (83).

10. Neglected dissipation. In the next four sections we shall discuss primarily the general characteristics of the steady-state solution for a line whose resistance and leakage are negligible. It may be argued that such a case does not occur in practice and that, therefore, its discussion is a waste of time.¹ The reader should bear in mind, however, that the performance of an actual line is best understood in the light of a modification of the behavior of the corresponding dissipationless line. In other words, the effect of resistance and leakage can be appreciated only after some familiarity has been gained with a system in which these parameters are absent. Only in the dissipationless case do some of the most interesting features concerning the steady-state line performance appear in their true light. The introduction of the parameters R and G , if sufficient in magnitude, may entirely obscure some

¹ In this connection it should be pointed out that high-frequency transmission lines used for feeding antennae are quite commonly treated by neglecting dissipation. See, for example, Hans Roder, "Graphical Methods for Problems Involving Radio-Frequency Transmission Lines," I.R.E. Proc., Vol. 21, No. 2, Feb., 1933, p. 290.

of the line's most peculiar properties which the student should not fail to appreciate. The effect of these parameters will, therefore, be taken up subsequently.

One of the first points of interest in a line without loss concerns itself with the propagation function (43a). Here we see that for $R = G = 0$ we get

$$\alpha = j\omega\sqrt{LC}, \quad (43b)$$

so that

$$\left. \begin{aligned} \alpha_1 &= 0 \\ \alpha_2 &= \omega\sqrt{LC} \end{aligned} \right\}. \quad (43c)$$

The component waves are, therefore, not attenuated as they travel. By (80a) the velocity of these waves is

$$v = \frac{\omega}{\alpha_2} = \frac{1}{\sqrt{LC}}, \quad (80b)$$

which, according to the discussion given in the first chapter, equals the velocity of light for open lines. The important point, however, is that this velocity is independent of frequency. This means that in the dissipationless case all frequencies have the same phase velocity.

The characteristic impedance according to (53) also becomes particularly simple for $R = G = 0$, namely

$$Z_0 = \sqrt{\frac{L}{C}}, \quad (53a)$$

which is a constant independent of frequency and hence is in effect a pure resistance. Thus for this case it is possible to terminate the line in its characteristic impedance. If this were done we would have by (74), (75), and (73a), assuming a pure resistance for Z_S and taking E_g as a reference

$$\left. \begin{aligned} e(x, t) &= \frac{Z_0 E_g}{Z_S + Z_0} \cos \omega(t - x\sqrt{LC}) \\ i(x, t) &= \frac{E_g}{Z_S + Z_0} \cos \omega(t - x\sqrt{LC}) \end{aligned} \right\}, \quad (73b)$$

where (43c) has been taken into account. This result makes it clear that, when a dissipationless line is properly terminated, it becomes a distortionless transmission system for the reason that neither the amplitude nor phase of the voltage or current is a function of frequency! This means that, when a complex periodic voltage is impressed, all its harmonic components will receive the same amplitude and phase changes during the process of transmission, so that the periodic function

at the receiving end will be an exact replica of that which was impressed. Later we shall see that transient functions may also be looked upon as the result of an infinite number of steady harmonic functions linearly superposed. Hence this distortionless property of the dissipationless line (when properly terminated) applies to all classes of impressed forces which may occur in practice.

The student should note well that the distortionless property of the dissipationless line does not follow unless the line is properly terminated, i.e., terminated in its characteristic impedance (53a). In order to convince himself of this he should introduce the relations (43c) into the general expressions (60a) and (61a), or their equivalents, and then form the instantaneous voltage and current expressions. He will find that unless $r_R = 0$, the resulting system will still possess distortion even though the propagation function is ideal. In general a line is used for communication in both directions, so that it will be necessary to properly terminate at both ends, since the sending and receiving ends continually interchange rôles. Proper termination is also essential from the standpoint of suppressing reflections which might otherwise cause echoes as well as interference with neighboring circuits.

It is also interesting to study the current and voltage relationships on the dissipationless line. Since these, however, involve the terminal impedances, we must first decide upon the nature of the latter before we can proceed with a more detailed discussion. Here two cases are of special interest. One is for pure reactive terminal impedances, i.e., for the case where the entire system, inclusive of its terminations, is dissipationless. The other is for pure resistance terminations associated with the dissipationless line. These we shall take up in the order named.

When Z_S and Z_R are pure reactances they are pure imaginary functions of frequency. In order to simplify the equations, let us write

$$\left. \begin{aligned} \frac{Z_S}{Z_0} &= jk_S \\ \frac{Z_R}{Z_0} &= jk_R \end{aligned} \right\} \quad (84)$$

Then we have for the reflection coefficients (62) and (63)

$$r_S = \frac{jk_S - 1}{jk_S + 1}, \quad (62a)$$

and

$$r_R = \frac{jk_R - 1}{jk_R + 1}. \quad (63a)$$

But these are complex numbers whose magnitudes are unity for all

values of k_S or k_R , and hence for all frequencies. The reflection coefficients are, therefore, pure rotational operators. If in this case we let

$$\left. \begin{aligned} r_S^{1/2} &= e^{j\vartheta_S} \\ r_R^{1/2} &= e^{j\vartheta_R} \end{aligned} \right\}, \quad (85)$$

then

$$\left. \begin{aligned} \vartheta_S &= \cot^{-1} k_S \\ \vartheta_R &= \cot^{-1} k_R \end{aligned} \right\}, \quad (85a)$$

and from (83) we see that

$$\left. \begin{aligned} \rho_S &= j\vartheta_S \\ \rho_R &= j\vartheta_R \end{aligned} \right\} \quad (83a)$$

so that the forms (60e) and (61e) for the complex voltage and current amplitudes become

$$E(x) = \frac{-E_g \cos [\alpha_2(l-x) - \vartheta_R]}{\sqrt{1 + k_S^2} \sin (\alpha_2 l - \vartheta_S - \vartheta_R)}, \quad (60f)$$

and

$$I(x) = \frac{-jE_g \sin [\alpha_2(l-x) - \vartheta_R]}{Z_0 \sqrt{1 + k_S^2} \sin (\alpha_2 l - \vartheta_S - \vartheta_R)}. \quad (61f)$$

For the exponential forms we see that by treating (64) in a similar manner to (82) we have

$$E^+ = \frac{-E_g e^{-j\vartheta_R}}{2\sqrt{1 + k_S^2} \sin (\alpha_2 l - \vartheta_S - \vartheta_R)}, \quad (64b)$$

and

$$E^- = E^+ e^{2j\vartheta_R}. \quad (64c)$$

The incident and reflected wave amplitudes for the current are these divided by $\pm Z_0$ respectively. Note that, if E_g is taken as a reference vector, E^+ and E^- form a conjugate pair, while I^- is the negative conjugate of I^+ . The exponential forms (66) become

$$\left. \begin{aligned} E(x) &= E^+ e^{j\alpha_2(l-x)} + E^- e^{-j\alpha_2(l-x)} \\ I(x) &= I^+ e^{j\alpha_2(l-x)} + I^- e^{-j\alpha_2(l-x)} \end{aligned} \right\}. \quad (66c)$$

Since these are in the form of the sum and difference of conjugate pairs respectively, their values equal 2 times the real part and $2j$ times the imaginary part of the first terms. The student may carry this out as an exercise and thus verify the results (60f) and (61f). He should note that in this case, where the entire system is dissipationless, the complex voltage and current amplitudes (60f) and (61f) are in quadrature at all points along the line, and also that, when E_g is taken

as a reference, E is purely real while I is a pure imaginary. Thus the power product is zero at all points in the system, as should be expected.

When the line itself is dissipationless, but terminated in pure resistances, the reflection coefficients (62) and (63) are real numbers whose values lie between plus and minus one, the first occurring for an open end and the second for a short-circuited end. The propagation function is, of course, pure imaginary. With these things in mind, the student can easily write down the expressions for this case using any of the above forms for the general one.

11. Standing waves. An interesting phenomenon occurs in the complete dissipationless system discussed in the previous section. Using (60f) and (61f) to form the instantaneous voltage and current according to (48), we have

$$e(x, t) = \frac{-E_0 \cos [\alpha_2(l - x) - \vartheta_R] \cdot \cos \omega t}{\sqrt{1 + k_S^2} \sin (\alpha_2 l - \vartheta_S - \vartheta_R)}, \quad (68a)$$

and

$$i(x, t) = \frac{E_0 \sin [\alpha_2(l - x) - \vartheta_R] \cdot \sin \omega t}{Z_0 \sqrt{1 + k_S^2} \sin (\alpha_2 l - \vartheta_S - \vartheta_R)}. \quad (69a)$$

These are harmonic functions of time whose amplitudes are harmonic functions of distance along the line. The voltage and current functions

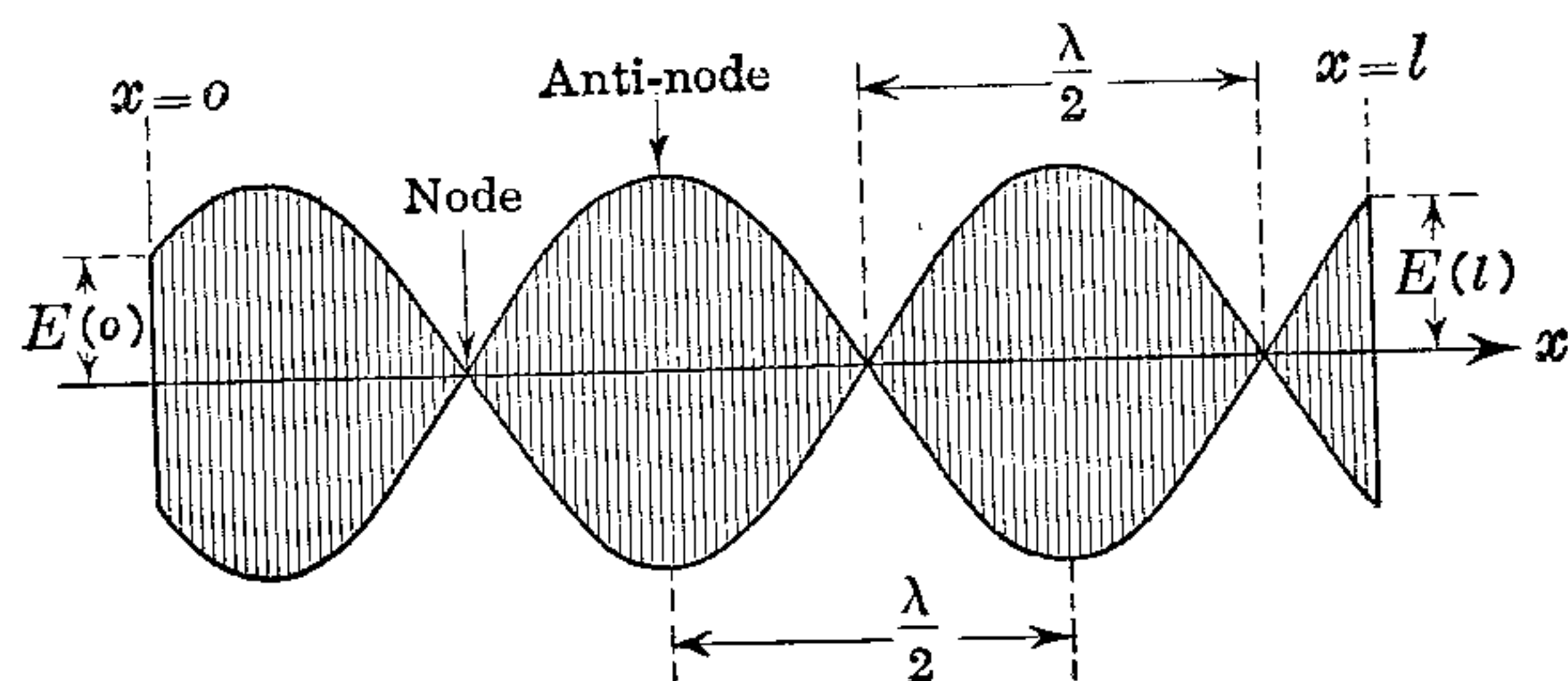


FIG. 5.—Diagrammatic representation of a standing wave.

are both in space and time quadrature. The time phase of each function is the same everywhere along the line. Thus the harmonic time variation is in unison throughout the line, and its amplitude varies harmonically along the line. There are points in the line where the amplitude is a maximum, and other points where the amplitude is zero at all times. The amplitude variation involved in this phenomenon is illustrated in Fig. 5. The harmonic envelope indicates the limits between which the synchronous harmonic time variation takes place. The result gives the appearance of a harmonic wave which is fixed in position, i.e., does not move. The phenomenon is therefore referred to as a **standing wave**. The point at which the amplitude is a maximum

is called an **anti-node**, while the zero points are called **nodes**. The distance between any pair of nodes or anti-nodes is obviously one-half a wave length. The phenomenon of standing waves thus places the significance of the wave length in evidence. This fact, together with the relation (78), which for the dissipationless case becomes

$$\lambda = \frac{2\pi}{\omega \sqrt{LC}} = \frac{\tau}{\sqrt{LC}} = \frac{v}{f}, \quad (78a)$$

is made use of in the laboratory for the measurement of high frequencies. A pair of large, well-insulated conductors are set up and coupled to a source of unknown frequency. A Geissler tube, whose length equals the spacing between conductors, is held so as to bridge these, and moved slowly along the line. By noting the points at which the tube extinguishes, the distance between two successive nodes is measured, thus determining λ . Assuming the velocity of light for v , the frequency can then be calculated. Such a system of parallel wires is called a *Lecher wire system*.¹

Note that the amplitude of an anti-node is determined primarily by the factor

$$\sin (\alpha_2 l - \vartheta_S - \vartheta_R)$$

in the denominator of either (68a) or (69a). If the length of the line and the terminal impedances are properly chosen, the argument of this sine function can be made equal to an integer number of π 's. The sine will then vanish, and the anti-nodes will theoretically become infinite. In any physical system, however, the incidental dissipation in the line and in the source will keep their magnitudes finite, but they will be a maximum for such an adjustment. In a Lecher wire system the ends are either left open or short-circuited, whichever is most convenient for a particular case, and the length of the line is made adjustable by means of a sliding mechanism of some kind so that the maximum condition just referred to can be obtained by trial. This adjustment obviously gives the maximum sensitivity for the device.

The student should note that when the far end is open the voltage has an anti-node there while the current has a node. With the far end short-circuited, the situation is the reverse. The reader may study various special cases as an exercise.

Before leaving the subject of standing waves, we wish to point out that this result in no way contradicts what has been said previously regarding the general wave character of the solution. The fact that, under special conditions, the net behavior has the appearance of a

¹ Lecher, Wied. Ann. 41, 1890, pp. 850-870.

standing wave does not mean that the incident and reflected wave components no longer move, or that one of them has disappeared and the other become stationary. Both are still there. The fact that the line is dissipationless means that neither wave attenuates. Owing to the non-dissipative terminal networks, the reflection coefficients have unit magnitude so that the phenomenon of reflection causes no change in amplitude but merely a change in phase. The incident and reflected waves, therefore, have the same amplitudes and do not attenuate. The phenomenon of two such unattenuated harmonic waves traveling in opposite directions with the same velocity gives rise to the standing wave effect. The phases of the waves with respect to a given point are of no consequence. The fact that their velocities are the same causes the net effect to be stationary. If one velocity were greater, then the net result would move in that direction.

12. Polar plots. Since $E(x)$ and $I(x)$ are vectors, the loci of their tips as functions of distance along the line for given terminal conditions are given by polar diagrams. In these diagrams the radius vectors represent the magnitudes and phase angles of the voltage and current at various points along the line. They are, therefore, particularly effective in lending pictorial significance to the line behavior.

An idea of the general character of the polar plots can best be obtained from a consideration of the forms (66) which represent the net voltage and current vectors in terms of the vector sums of incident and reflected vector-components. The individual polar plots of these component vectors are obviously logarithmic spirals. For example we have for the incident voltage vector

$$E^+ \cdot e^{\alpha(l-x)} = E^+ \cdot e^{\alpha_1(l-x)} \cdot e^{j\alpha_2(l-x)}.$$

Starting with $x = 0$ and proceeding toward $x = l$, this represents a vector which decays exponentially and rotates uniformly in the clockwise direction, the rates of decay and rotation being determined by α_1 and α_2 respectively. The tip of this vector traces a logarithmic spiral.

The reflected voltage vector

$$E^- \cdot e^{-\alpha(l-x)} = E^- \cdot e^{-\alpha_1(l-x)} \cdot e^{-j\alpha_2(l-x)}$$

may be represented in a similar manner. This vector, however, decays exponentially and rotates uniformly in a clockwise direction as x is allowed to decrease from $x = l$ to $x = 0$. At the end $x = l$ the incident vector is E^+ and the reflected vector is E^- . If the reflection coefficient r_R has a magnitude different from unity and also has an angle, the E^- will be different from E^+ both in magnitude and angle.

The net voltage at any point in the line is the vector sum of the vectors representing the incident and reflected waves at that point.

If the logarithmic spirals for these are superposed, and a sufficient number of points corresponding to arbitrarily chosen x -values marked on each, then the resultant vectors for these points may be found and a continuous curve drawn through their tips. This resultant polar plot will show the variation in magnitude and phase of the actual line voltage with respect to distance. Its shape is that of an elliptic spiral, although it may be difficult to recognize this if the attenuation is high.

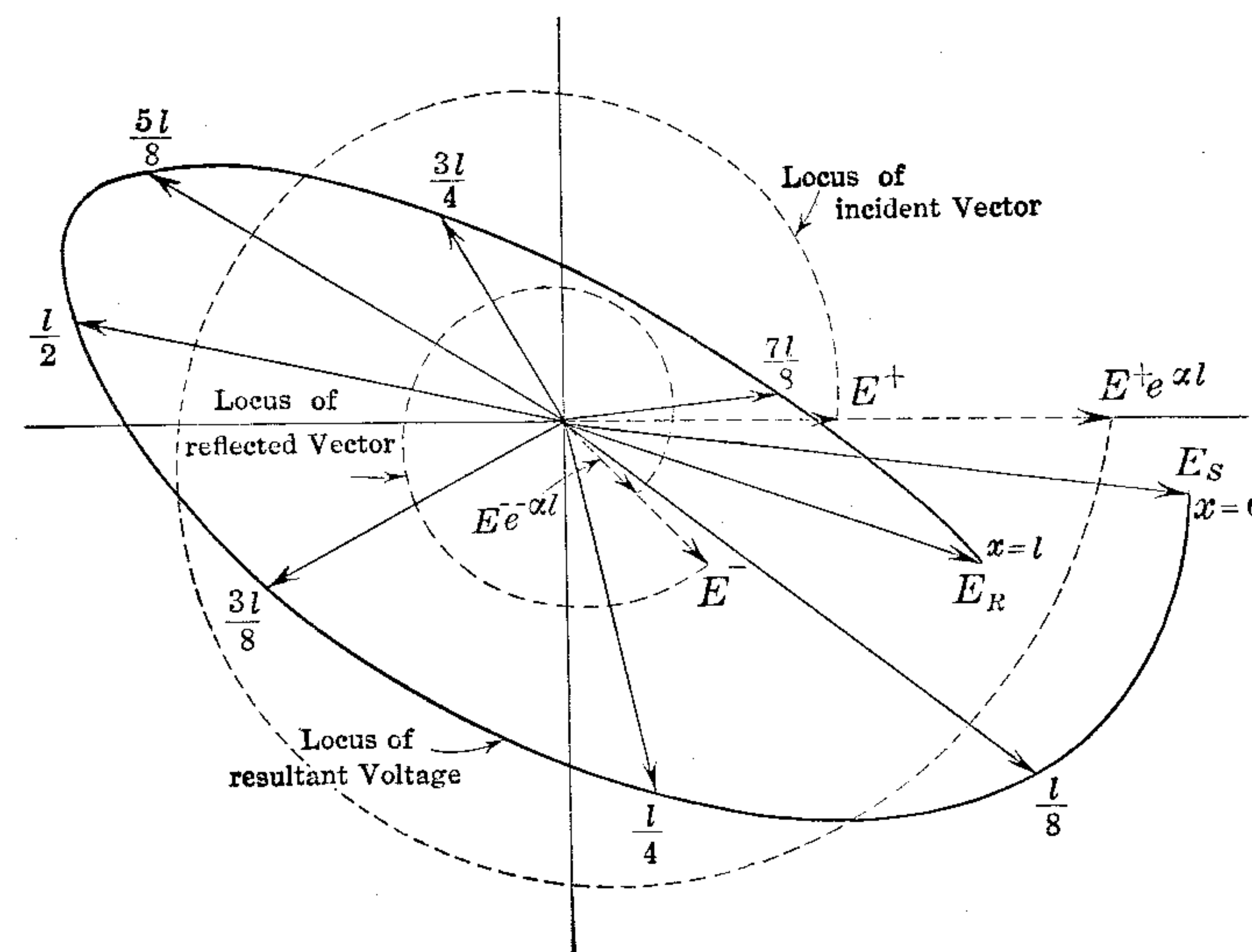


FIG. 6.—Polar plot showing the variation of the vector voltage along a dissipative line one wave length long.

The same procedure may be carried out for the current. Points may be marked on the resultant plots corresponding to fractions of the total line length measured from the sending end. If a sufficient number of points are thus indicated, the plots may be used to read graphically the magnitudes and phase angles of $E(x)$ and $I(x)$ at any desired point in the line. Fig. 6 illustrates the polar plot for voltage for the following arbitrarily chosen data

$$\left. \begin{aligned} l &= \lambda; \alpha_2 l = 2\pi \\ \alpha_1 &= 0.1104\alpha_2; e^{\alpha_1 l} = 2 \\ E^+ &\text{ taken as reference} \\ r_R &= 0.75 \left[-\frac{\pi}{4} \right] \end{aligned} \right\}.$$

Several points may be emphasized with regard to this figure. The student should note well that although the component vectors rotate

uniformly with distance, the resultant does not. In the figure the resultant voltage vectors are drawn for each eighth of a line length. It is quite apparent that the angles between the various vectors are not equal; nor is the total phase change in the net voltage 2π radians although the line is one wave length long. The incident and reflected components, on the other hand, rotate through $\pi/4$ radians for each interval $l/8$. It is also clear from the figure that the reflected vector has its greatest effect in the vicinity of the receiving end of the line. If the figure were carried out for an additional wave length, the resultant locus would very nearly coincide with that of the incident vector in the vicinity of the sending end. When the line is terminated in its characteristic impedance, the net voltage locus becomes the logarithmic spiral representing the incident component.

When line losses are neglected, the situation regarding polar plots becomes very much simpler. The fact that this case is of practical as well as theoretical importance has already been pointed out. When the attenuation on the line is zero, the loci for the incident and reflected vector-components become circles, and the resultant locus is an ellipse. This may be shown in the following way:

Using the relations (66) and (64a) we may write for the voltage in the non-dissipative case

$$E(x) = E^+(e^{j\alpha_2(l-x)} + r_R e^{-j\alpha_2(l-x)}). \quad (86)$$

Assuming an arbitrary terminal impedance Z_R , the reflection coefficient r_R will be complex. If we write

$$r_R = |r_R| e^{2j\vartheta_R}, \quad (87)$$

then (86) becomes

$$E(x) = E^+ \cdot e^{j\vartheta_R} (e^{j[\alpha_2(l-x) - \vartheta_R]} + |r_R| e^{-j[\alpha_2(l-x) - \vartheta_R]}). \quad (88)$$

Since $E^+ \cdot e^{j\vartheta_R}$ is a factor which is independent of x , it will be more convenient and just as useful to discuss the ratio

$$\frac{E(x)}{E^+} \cdot e^{-j\vartheta_R} = \mathcal{E} = \mathcal{E}_1 + j\mathcal{E}_2 \quad (89)$$

in place of the voltage E itself. The latter is simply rotated and stretched according to the angle and magnitude of $E^+ \cdot e^{j\vartheta_R}$. Since

$$E(l) = E_R = E^+(1 + r_R), \quad (90)$$

the relation (88) may also be expressed in terms of the voltage at the receiving end instead of E^+ .

Combining (88) and (89) and writing sines and cosines for the exponentials, we get by equating reals and imaginaries

$$\left. \begin{aligned} \mathcal{E}_1 &= (1 + |r_R|) \cos [\alpha_2(l-x) - \vartheta_R] \\ \mathcal{E}_2 &= (1 - |r_R|) \sin [\alpha_2(l-x) - \vartheta_R] \end{aligned} \right\} \quad (91)$$

Between these equations it is possible to eliminate either $|r_R|$ or x . In order to do the latter, we divide the first equation by $(1 + |r_R|)$, the second by $(1 - |r_R|)$, square each separately, and then add. We then have

$$\frac{\mathcal{E}_1^2}{(1 + |r_R|)^2} + \frac{\mathcal{E}_2^2}{(1 - |r_R|)^2} = 1, \quad (92)$$

which is the equation of an ellipse in rectangular coordinates with the semi-major axis $(1 + |r_R|)$, and semi-minor axis $(1 - |r_R|)$. Every point on this ellipse corresponds to a terminal impedance giving rise to the same $|r_R|$, but to a definite point in the line. That is, various points on the contour of the ellipse represent various points along the line. We have yet to show how a point on the ellipse may be located to correspond to a given point in the line. This correlation is given by the following.

Suppose we eliminate $|r_R|$ from (91) by dividing by the trigonometric functions and then adding, thus

$$\frac{\mathcal{E}_1}{\cos [\alpha_2(l-x) - \vartheta_R]} + \frac{\mathcal{E}_2}{\sin [\alpha_2(l-x) - \vartheta_R]} = 2. \quad (93)$$

This is the parametric equation of a straight line in cartesian coordinates. It is a straight line with the intercepts $2\cos [\alpha_2(l-x) - \vartheta_R]$ on the horizontal, and $2\sin [\alpha_2(l-x) - \vartheta_R]$ on the vertical axis. The constant distance between these intercepts is 2. The angle $[\alpha_2(l-x) - \vartheta_R]$ is that included between the line and the horizontal axis. This line may be looked upon as a locus on which $|r_R|$ varies from point to point while x and ϑ_R remain constant. Thus if the ellipse (92) and the line (93) are superposed, their intersection determines the vector $\mathcal{E}$ corresponding to the values x , ϑ_R , and $|r_R|$ for which the loci are drawn.

It is well known in geometry, however, that a fixed point on the line (93) will trace an ellipse when the angle of the sine and cosine functions is allowed to pass from zero to 2π , the sum of the major and minor axes of this ellipse being 2. This is none other than the ellipse (92) when the fixed point is so located that its distances from the intercepts with the horizontal and vertical axes are $(1 - |r_R|)$ and $(1 + |r_R|)$ respectively. Thus the line (93) may be used to generate the desired locus (92); and the distance x is thus correlated with each point in this

locus by the angle $[\alpha_2(l-x) - \vartheta_R]$. This is illustrated in Fig. 7. The length PQ is constant and equal to 2. The point R which is determined from $|r_R|$ traces the ellipse as x varies, i.e., as the line is rotated so as to keep P and Q on the vertical and horizontal axes respectively.

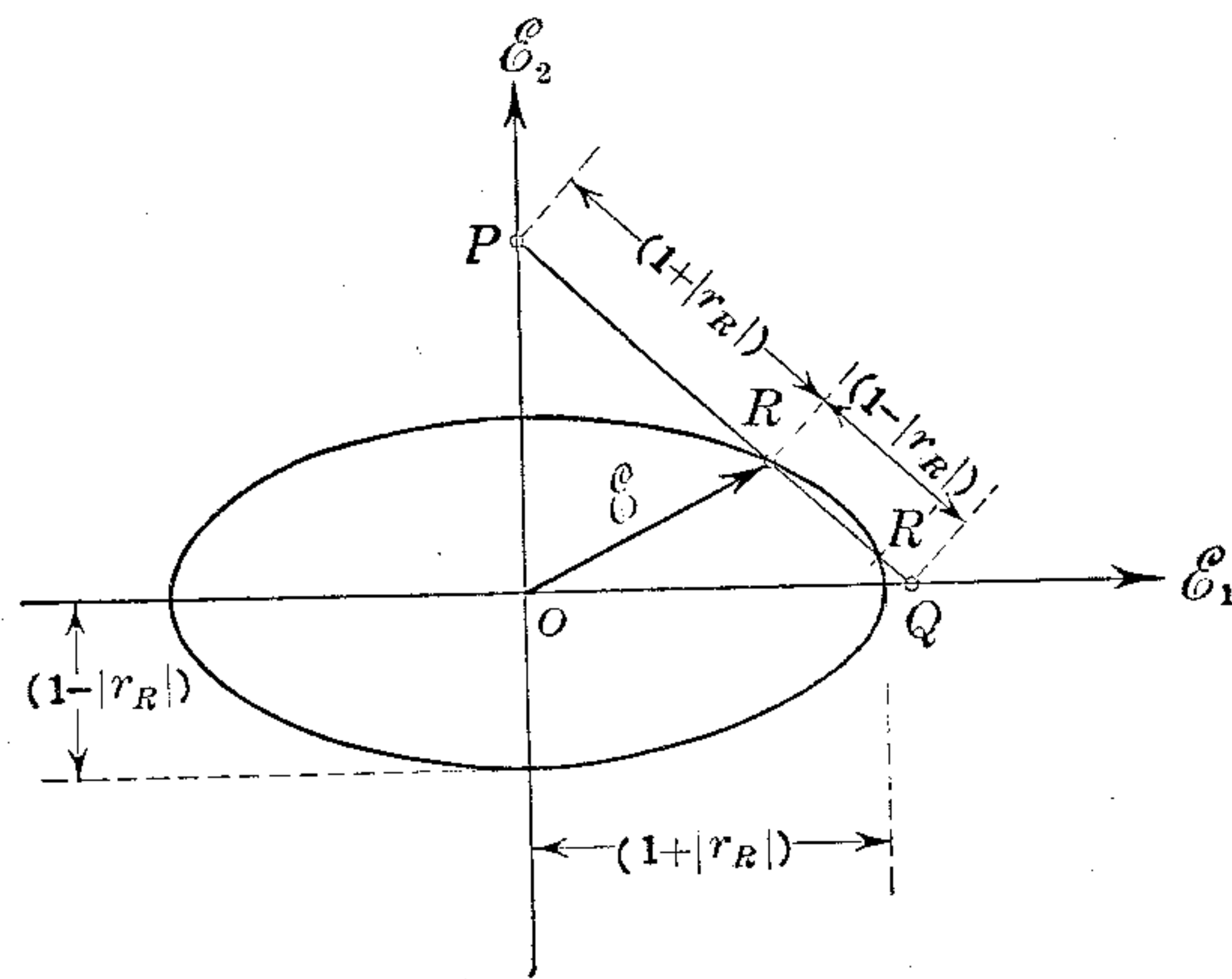


FIG. 7.—Elliptic polar plot of the voltage variation along a non-dissipative line with arbitrary terminal impedance.

and hence the largest value which $|r_R|$ can have is unity. The range zero-to-unity for $|r_R|$, therefore, exhausts all the possibilities for the case under discussion. This means that the point R will always lie between P and Q . In the limiting case $|r_R| = 1$, which occurs for either the open or short-circuited line, R coincides with Q , and the ellipse degenerates into a double straight line lying in the horizontal axis. The same occurs for any pure reactive termination. For $Z_R = Z_0$ the reflection coefficient vanishes, and R lies midway between P and Q . The resulting locus is a circle, as is to be expected. Fig. 8 shows a family of loci for various values of terminal impedance. The horizontal axis between mm' is the degenerate ellipse.

When the line PQ is superposed upon the family of loci of Fig. 8, it may seem confusing to pick the proper intersections with the variou

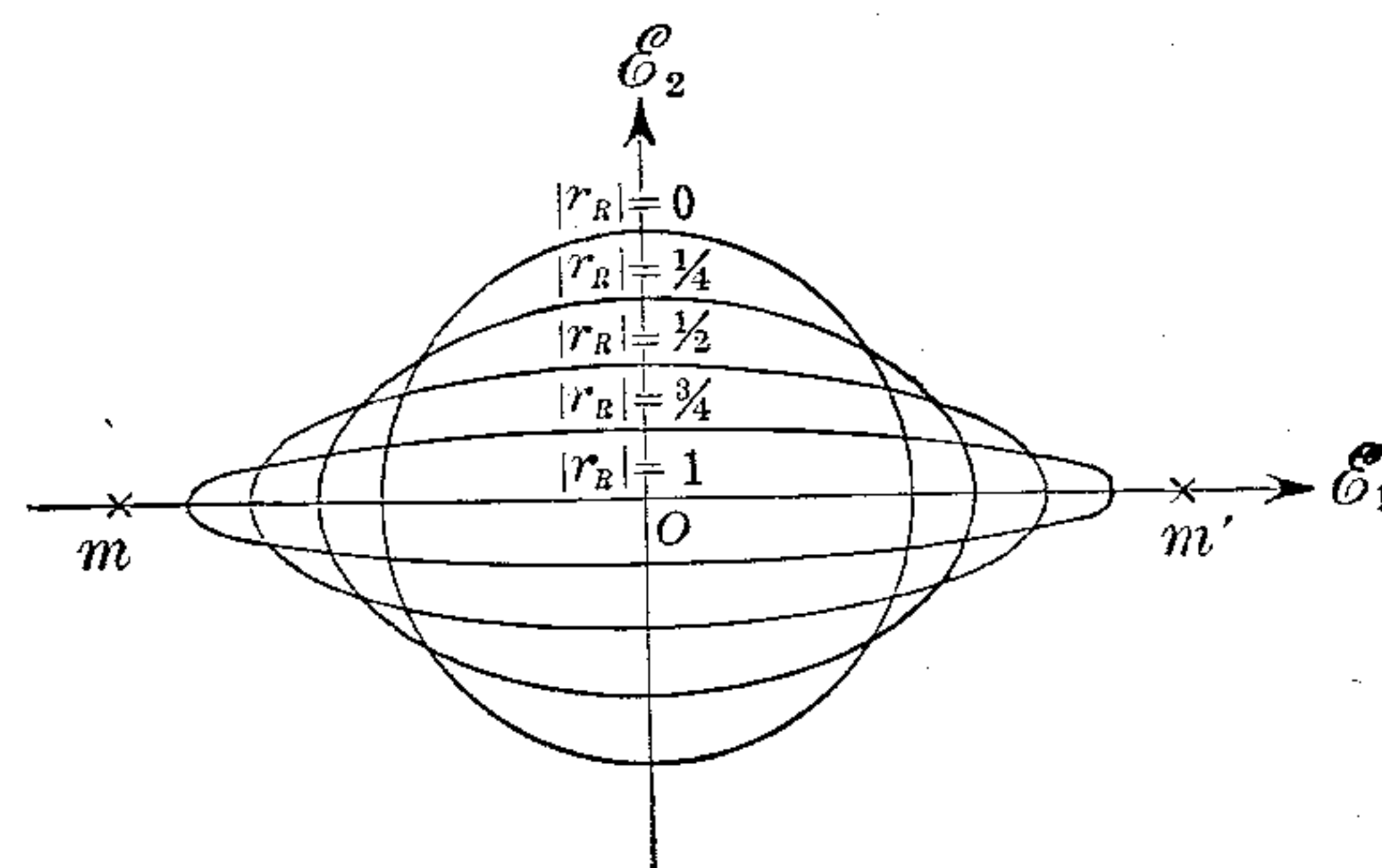


FIG. 8.—Elliptic voltage loci for various values of terminal impedance.

This method of correlation makes the geometric construction very simple. In fact, to find $\mathcal{E}$ for a given x and r_R , it is not necessary to draw the ellipse at all. The line PQ , two units in length, is oriented according to the given x and ϑ_R . The point R is located by making PR equal to $(1 + |r_R|)$. The vector OR is then the desired $\mathcal{E}$.

When line losses are neglected, Z_0 is real,

ellipses owing to the fact that the line will have a second intersection with some of these, as for example at the point R' in Fig. 7. Such an intersection must be disregarded on the ground that the point in question does not generate the corresponding ellipse, as may easily be seen.

The loci giving the current variation along the line may be obtained in the same fashion. Using (65), (65a), and (66) we have

$$I(x) = \frac{E^+ e^{j\vartheta_R}}{Z_0} (e^{j[\alpha_2(l-x) - \vartheta_R]} - |r_R| e^{-j[\alpha_2(l-x) - \vartheta_R]}) \quad (94)$$

in which (87) is again substituted for r_R . Defining a vector $\mathfrak{I}$ by the relation

$$\frac{Z_0 I}{E^+ e^{j\vartheta_R}} = \mathfrak{I} = \mathfrak{I}_1 + j\mathfrak{I}_2, \quad (95)$$

we have

$$\begin{aligned} \mathfrak{I}_1 &= (1 - |r_R|) \cos [\alpha_2(l-x) - \vartheta_R] \\ \mathfrak{I}_2 &= (1 + |r_R|) \sin [\alpha_2(l-x) - \vartheta_R] \end{aligned} \quad (96)$$

from which we obtain the following equations for the ellipse and straight line

$$\frac{\mathfrak{I}_1^2}{(1 - |r_R|)^2} + \frac{\mathfrak{I}_2^2}{(1 + |r_R|)^2} = 1 \quad (97)$$

$$\frac{\mathfrak{I}_1}{\cos [\alpha_2(l-x) - \vartheta_R]} + \frac{\mathfrak{I}_2}{\sin [\alpha_2(l-x) - \vartheta_R]} = 2. \quad (98)$$

Comparing these with (92) and (93) we see that the corresponding current and voltage ellipses have their major and minor axes interchanged, while the straight lines are the same for both voltage and current. Whereas for the voltage ellipse the major axis is horizontal, it coincides with the vertical axis for the current ellipse. The method of constructing the current loci needs no additional comment.¹

These polar plots show that whenever the line is not terminated in its characteristic impedance the voltage and current magnitudes alter-

¹ It is interesting to note that, in the case of real values of Z_R (pure resistance termination), the above relations become particularly simple. We then have for the voltage and current vectors in terms of the receiving-end voltage and current

$$\begin{aligned} E(x) &= E_R \cos \alpha_2(l-x) + jE_R \cdot \frac{Z_0}{Z_R} \sin \alpha_2(l-x) \\ I(x) &= I_R \cos \alpha_2(l-x) + jI_R \cdot \frac{Z_R}{Z_0} \sin \alpha_2(l-x) \end{aligned}$$

Taking E_R as a reference, the first of these represents an ellipse with E_R as its semi-axis in the horizontal direction and $E_R \cdot \frac{Z_0}{Z_R}$ in the vertical direction. A similar situation holds for the current ellipse with I_R as reference.

nately pass through maximum and minimum values as the length of the line is traversed, the magnitude of the variation being larger for larger departures from proper termination. When line losses are neglected, the maxima and minima occur at quarter-wave-length intervals. The fact that the major axes of the ellipse diagrams for voltage and current are in quadrature means that the respective maxima and minima replace each other; i.e., the voltage has a maximum where the current has a minimum, and vice versa.

This phenomenon of the appearance of maxima and minima, which is accentuated in standing waves, is called the **Ferranti effect**. The presence of dissipation, if sufficient in magnitude, will tend to suppress the effect more or less completely.

13. Odd and even quarter-wave-length lines. When, in the non-dissipative case, the line length equals either an odd or even multiple of a quarter wave length, some peculiar characteristics appear.

Referring to the general expressions (60a) and (61a) for the complex voltage and current amplitudes, we obtain the following expressions for ratios of receiving-to-sending-end quantities for a non-dissipative line with arbitrary terminal impedances

$$\frac{E(l)}{E(0)} = \frac{1 + r_R}{e^{j\alpha_2 l} + r_R e^{-j\alpha_2 l}}, \quad (99)$$

and

$$\frac{I(l)}{E(0)} = \frac{1 - r_R}{Z_0(e^{j\alpha_2 l} + r_R e^{-j\alpha_2 l})}. \quad (100)$$

For

$$l = \frac{\nu\lambda}{4} = \frac{\nu\pi}{2\alpha_2}; (\nu = 1, 3, 5, \dots), \quad (101)$$

we have

$$e^{j\alpha_2 l} + r_R e^{-j\alpha_2 l} = e^{\frac{j\nu\pi}{2}} (1 + r_R e^{-j\nu\pi}) = j(1 - r_R)(-1)^{\frac{\nu-1}{2}}. \quad (102)$$

Noting that by (63)

$$\frac{1 + r_R}{1 - r_R} = \frac{Z_R}{Z_0}, \quad (103)$$

we get for this case

$$\frac{E(l)}{E(0)} = j \frac{Z_R}{Z_0} \cdot (-1)^{\frac{\nu+1}{2}}, \quad (99a)$$

and

$$\frac{I(l)}{E(0)} = j \frac{(-1)^{\frac{\nu+1}{2}}}{Z_0}. \quad (100a)$$

On the other hand for

$$l = \frac{\nu\lambda}{4} = \frac{\nu\pi}{2\alpha_2}; (\nu = 0, 2, 4, \dots), \quad (101a)$$

we have

$$e^{j\alpha_2 l} + r_R e^{-j\alpha_2 l} = (1 + r_R)(-1)^{\frac{\nu}{2}}, \quad (102a)$$

and hence

$$\frac{E(l)}{E(0)} = (-1)^{\frac{\nu+2}{2}}, \quad (99b)$$

and

$$\frac{I(l)}{E(0)} = \frac{Z_0}{Z_R}(-1)^{\frac{\nu+2}{2}}. \quad (100b)$$

For the line lengths (101), for which (99a) and (100a) follow, we see that if we consider the sending-end voltage $E(0)$ constant, the received voltage becomes a linear function of the load impedance Z_R , while the load current becomes independent of the load impedance, i.e., a constant! Thus the line will virtually transform a constant-voltage system into a constant-current system. For the line lengths (101a), on the other hand, the received voltage becomes independent of the load impedance, while the current varies inversely as that impedance. Here the effect of the line is nil except in so far as it may affect the sign of the voltage and current. In practice these conditions can, of course, be only approximated. So far as communication work is concerned, such conditions are neither desirable nor in evidence. We merely mention them here for the sake of general interest. Properties such as these are also exhibited by certain lumped-constant networks at properly chosen frequencies. These were first described by Boucherot and are sometimes named after him.

14. Effect of attenuation. In actual telephone lines and cables, the variation in voltage and current along the line is quite materially influenced by the presence of resistance and leakage, so that the discussion in the last four sections can serve only to indicate the general trend of these functions. In order to illustrate this, let us consider the following parameter values per mile which are fairly representative of an open telephone line consisting of a pair of 104 mil copper wires with 18-inch spacing.¹ Here

$$R = 10; L = 0.004; G = 10^{-6}; C = 0.008 \cdot 10^{-6}.$$

¹ The 18-inch spacing is for a so-called pole-pair, i.e., that pair of wires which straddles the pole. The parameters for a non-pole-pair do not differ appreciably from these.

The frequency at which these values hold corresponds to $\omega = 5,000$ radians per second. Using these in (43a) we find

$$\alpha = 0.00724 + j 0.029 \quad (104)$$

i.e.,

$$\left. \begin{array}{l} \alpha_1 = 0.00724 \\ \alpha_2 = 0.029 \end{array} \right\} \quad (104a)$$

For the characteristic impedance we have by (53)

$$Z_0 = 748 \angle 12.5^\circ. \quad (105)$$

Using (78) and (104a), the wave length becomes

$$\lambda = \frac{2\pi}{0.029} = 216.5 \text{ miles}, \quad (106)$$

while for the phase velocity we get

$$v = \frac{\omega}{\alpha_2} = \frac{5000}{0.029} = 172,400 \text{ miles per second} \quad (107)$$

which is slightly less than the velocity of light.

It is interesting to compare these values with those which would result if the resistance and leakage were zero. Strictly speaking the inductance would become somewhat less if the resistance were zero because then the skin effect would be complete; but we shall neglect this small change. Then we get

$$\left. \begin{array}{l} \alpha = j 0.0283 \\ Z_0 = 707 \\ \lambda = 222 \\ v = 176,600 \end{array} \right\} \quad (108)$$

Except for the fact that the attenuation is zero, these values do not differ very much from those for the actual line. The characteristic impedance shows perhaps the most significant change, but this will ordinarily not affect the line behavior to a great extent. This may be seen if we examine the effect which a change in Z_0 has upon the reflection coefficient r_R . Suppose we let

$$Z_0' = Z_0(1 + \Delta), \quad (109)$$

where Δ denotes the decimal change in Z_0 . Then the new reflection coefficient becomes

$$r_R' = \frac{Z_R - Z_0(1 + \Delta)}{Z_R + Z_0(1 + \Delta)}. \quad (110)$$

Expanding and retaining only linear terms, we find for this the following approximate expression

$$r_R' = r_R - \frac{2 Z_R Z_0 \Delta}{(Z_R + Z_0)^2} = r_R - \Delta', \quad (110a)$$

where r_R is the value for $\Delta = 0$. The effect upon the reflection coefficient is expressed by the second term in (110a). This is shown plotted versus Z_R in Fig. 9, from which we see that the change in the reflection coefficient has a maximum value at $Z_R = Z_0$, and that this maximum change equals half the decimal change in Z_0 . If the latter were, say, 0.1, then the maximum variation that this could produce in r_R would be 0.05. Note, however, that this is the actual change in r_R , not the percentage change which would be infinite for $Z_R = Z_0$

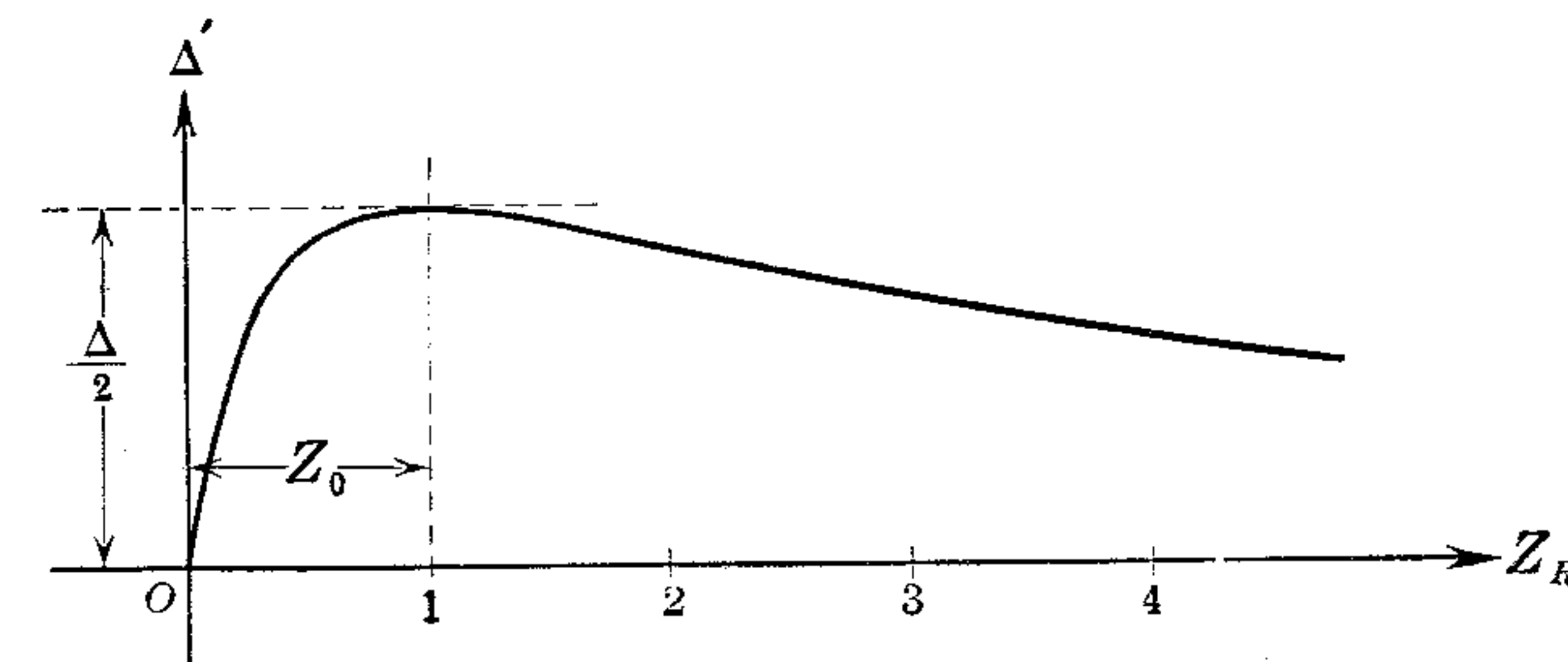


FIG. 9.—Change in the value of the reflection coefficient r_R as a function of the terminal impedance.

because then $r_R = 0$. For values of Z_R not in the vicinity of Z_0 , the effect upon r_R is much less and hence for most purposes negligible.

Thus the outstanding effect in the dissipative case, at least so far as the behavior at a single frequency is concerned,¹ is the presence of attenuation. In our numerical example, (104a) gives the attenuation per mile. For a wave length the attenuation factor would be by (106)

$$216.5 \times 0.00724 = 1.568 \text{ napiers}$$

so that the ratio of the magnitudes of a single wave component at intervals of a wave length would be

$$e^{1.568} = 4.8,$$

i.e., the magnitude of either the incident or reflected wave decreases almost 80 per cent as it travels a distance of one wave length or 216.5 miles. Fig. 10 shows the variation in the net voltage over 111 miles (one-half wave length for the corresponding non-dissipative case) for a terminal resistance of 2,121 ohms (this makes $r_R = 0.5$ for neglected

¹ When the behavior is studied as a function of frequency, other very significant effects will be found.

R and G). Since this plot is drawn for the ratio E/E^+ , the actual voltage is found by multiplying by E^+ which must be determined from (64). In the same figure is also shown the ellipse which is obtained for the corresponding non-dissipative case. A comparison of this ellipse and the polar curve for the actual line will give the reader some idea of the effect which the attenuation has upon the line behavior. The effect upon the phase function is evidenced by the fact that the vectors for $x = 0$ and $x = l$ are not exactly horizontal. The deviation, however, is not very great.

The student should draw the corresponding polar plot for the current as well as those for a purely reactive terminal impedance.

15. General circuit parameters. Very often in communication work we are not so much interested in the variation of voltage and current

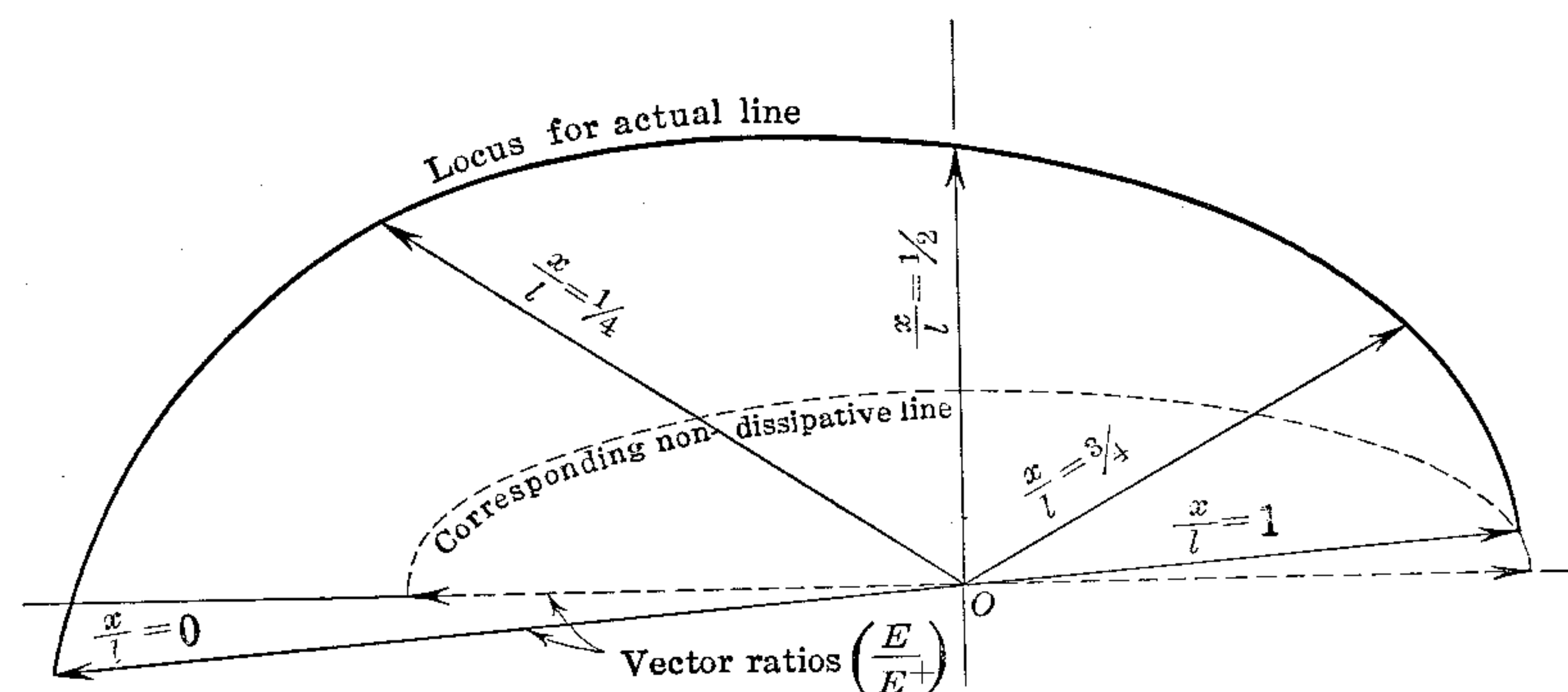


FIG. 10.—Polar plots showing the relative effect of dissipation.

along the line as we are in the relations between the sending- and receiving-end quantities alone. Hence it is convenient to have mathematical expressions which involve only such quantities. These may most conveniently be obtained from the hyperbolic forms (60d) and (61d). For example, if in the first of these we let $x = l$, we find

$$E_R = \frac{Z_0 Z_R E_S}{Z_0(Z_S + Z_R) \cosh \alpha l + (Z_S Z_R + Z_0^2) \sinh \alpha l}; \quad (111)$$

and if we use this relation together with

$$I_R Z_R = E_R, \quad (112)$$

the equations (60d) and (61d) may be expressed in terms of E_R and I_R , thus

$$\left. \begin{aligned} E(x) &= E_R \cosh \alpha(l-x) + Z_0 I_R \sinh \alpha(l-x) \\ I(x) &= \frac{E_R}{Z_0} \sinh \alpha(l-x) + I_R \cosh \alpha(l-x) \end{aligned} \right\} \quad (113)$$

In particular, for $x = 0$ we have

$$\left. \begin{aligned} E_S &= E_R \cosh \alpha l + Z_0 I_R \sinh \alpha l \\ I_S &= \frac{E_R}{Z_0} \sinh \alpha l + I_R \cosh \alpha l \end{aligned} \right\} \quad (114)$$

which express the sending- in terms of the receiving-end quantities directly. In order to do the reverse, we have merely to solve this set simultaneously for E_R and I_R in terms of E_S and I_S . This gives

$$\left. \begin{aligned} E_R &= E_S \cosh \alpha l - Z_0 I_S \sinh \alpha l \\ I_R &= -\frac{E_S}{Z_0} \sinh \alpha l + I_S \cosh \alpha l \end{aligned} \right\} \quad (115)$$

When applying the results (114) and (115), the reader should recall the conventions as to positive directions for voltage and current, otherwise erroneous conclusions may be drawn. Fig. 11 illustrates these conventions schematically.

Positive flow of current is from left to right, and a positive voltage is one which rises from the datum plane to the conductor in the single-line diagram. The fact that the currents have this left-to-right direction causes the minus signs which appear in (115) but which are absent in (114). If the current directions are reversed, the signs in (115) will all become positive, but corresponding negative signs will appear in (114). Except for this matter of signs, the relations (114) and (115) form a symmetrical set; i.e., we may obtain one pair from the other by simply interchanging the subscripts S and R . This is to be expected from the fact that the line is a symmetrical system with respect to its two ends.

The coefficients in the equations (114) are called the **general circuit parameters** and are usually denoted by

$$\left. \begin{aligned} \mathfrak{A} &= \cosh \alpha l \\ \mathfrak{B} &= Z_0 \sinh \alpha l \\ \mathfrak{C} &= \frac{1}{Z_0} \sinh \alpha l \\ \mathfrak{D} &= \cosh \alpha l \end{aligned} \right\} \quad (116)$$

In terms of these, we have

$$\left. \begin{aligned} E_S &= \mathfrak{A}E_R + \mathfrak{B}I_R \\ I_S &= \mathfrak{C}E_R + \mathfrak{D}I_R \end{aligned} \right\}, \quad (114a)$$

and

$$\left. \begin{aligned} E_R &= \mathfrak{D}E_S - \mathfrak{B}I_S \\ I_R &= -\mathfrak{C}E_S + \mathfrak{A}I_S \end{aligned} \right\}. \quad (115a)$$

Mathematically, the pair of relations (114a) is spoken of as a linear transformation of the quantities E_R and I_R into E_S and I_S . The following arrangement of coefficients

$$\begin{vmatrix} \mathfrak{A} & \mathfrak{B} \\ \mathfrak{C} & \mathfrak{D} \end{vmatrix} \quad (117)$$

is called the *transformation matrix*, and

$$\begin{vmatrix} \mathfrak{A} & \mathfrak{B} \\ \mathfrak{C} & \mathfrak{D} \end{vmatrix} \quad (118)$$

the *determinant of the transformation*. The relations (115a) represent the inverse transformation, and

$$\begin{vmatrix} \mathfrak{D} & -\mathfrak{B} \\ -\mathfrak{C} & \mathfrak{A} \end{vmatrix} \quad (117a)$$

the *inverse matrix* to (117). In Chapter IV we shall show how the behavior of any linear four-terminal network may be analyzed in terms of its transformation matrix. The interconnection of several systems may be most effectively treated in this way.

Note that the transformation determinant (118) has the value

$$\mathfrak{A}\mathfrak{D} - \mathfrak{B}\mathfrak{C} = 1 \quad (119)$$

as may be easily verified by the use of (116). From the latter we also see that in this case

$$\mathfrak{A} = \mathfrak{D}. \quad (120)$$

Thus we see that, of the four general circuit parameters $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$, $\mathfrak{D}$, only two are independent. We shall find later that the relation (119) holds generally for all linear four-terminal networks, and that, in addition, (120) holds for all those which are symmetrical with respect to their input and output terminals.

16. Input impedance. It is sometimes necessary to determine the impedance which the line presents at various points with an arbitrary load impedance. This may easily be done in terms of either the exponential forms (60a) and (61a), or the hyperbolic forms (60e) and

(61e) involving the equivalent line angles (83). Thus we get for the impedance looking into the line at any point x

$$\left. \begin{aligned} \frac{E(x)}{I(x)} &= Z(x) = Z_0 \coth [\alpha(l-x) - \rho_R] \\ &= Z_0 \frac{e^{\alpha(l-x)} + r_R e^{-\alpha(l-x)}}{e^{\alpha(l-x)} - r_R e^{-\alpha(l-x)}} \end{aligned} \right\}. \quad (121)$$

For the impedance looking in at the sending end, we have merely to put $x = 0$. Here two special cases are of interest. Namely, for $Z_R = 0$ we have $r_R = -1$, and by the second relation (121)

$$Z(0) = Z_{sc} = Z_0 \tanh \alpha l, \quad (122)$$

while for $Z_R = \infty$, $r_R = +1$, so that

$$Z(0) = Z_{oc} = Z_0 \coth \alpha l. \quad (123)$$

Z_{oc} and Z_{sc} are called the *open-circuit* and *short-circuit* input impedances of the line respectively. In terms of these the propagation function and characteristic impedance may be expressed. By forming the product or quotient of (122) and (123) we find, respectively

$$Z_0 = \sqrt{Z_{oc} \cdot Z_{sc}}, \quad (124)$$

and

$$\alpha l = \tanh^{-1} \sqrt{\frac{Z_{sc}}{Z_{oc}}} = \frac{1}{2} \ln \left(\frac{\sqrt{\frac{Z_{oc}}{Z_{sc}}} + 1}{\sqrt{\frac{Z_{oc}}{Z_{sc}}} - 1} \right). \quad (125)$$

These relations may be used to determine Z_0 and α experimentally from measurements of Z_{oc} and Z_{sc} . It is rather interesting that the characteristic impedance is determined by the product, and the propagation function by the quotient, of these impedances.

The student should note that the line itself is completely characterized by two functions no matter what form the characterization may take. So far we have seen that the line performance is expressible in terms of Z_0 and α , or any two of the general circuit parameters discussed in the previous section, or in terms of Z_{oc} and Z_{sc} as is evident from (124) and (125). The fact that, in any case, *two* functions suffice to describe the line characteristics uniquely, is fundamental.

17. Voltage, current, and power ratios. Very often we are not so much interested in the actual values of voltage, current, or power received at the far end of a line as we are in the ratio of output to input or its inverse. This ratio as a function of frequency is a direct indication of the quality of a given transmission facility.

ATHERFORCE

Using the voltage equation (60a), we get for $x = l$

$$\frac{E_R}{E_g} = \frac{Z_0(1 + r_R)}{(Z_S + Z_0)(e^{\alpha l} - r_S r_R e^{-\alpha l})}. \quad (126)$$

For purposes of discussion, this may be put in a more effective form. Using (62) and (63) we find that

$$\frac{Z_0(1 + r_R)}{(Z_S + Z_0)} = \frac{Z_R}{Z_S + Z_R} \cdot (1 - r_S r_R), \quad (127)$$

so that we have

$$\frac{E_g}{E_R} = \frac{Z_S + Z_R}{Z_R} \cdot e^{\alpha l} \cdot (1 - r_S r_R)^{-1} \cdot (1 - r_S r_R e^{-2\alpha l}). \quad (128)$$

This is the desired ratio of input to output voltage. The reason for separating it into factors in the above manner is for convenience in discussion and calculation. The significance of the first factor, which involves only the sending- and receiving-end impedances, will be brought out shortly. The second or exponential factor is the most important

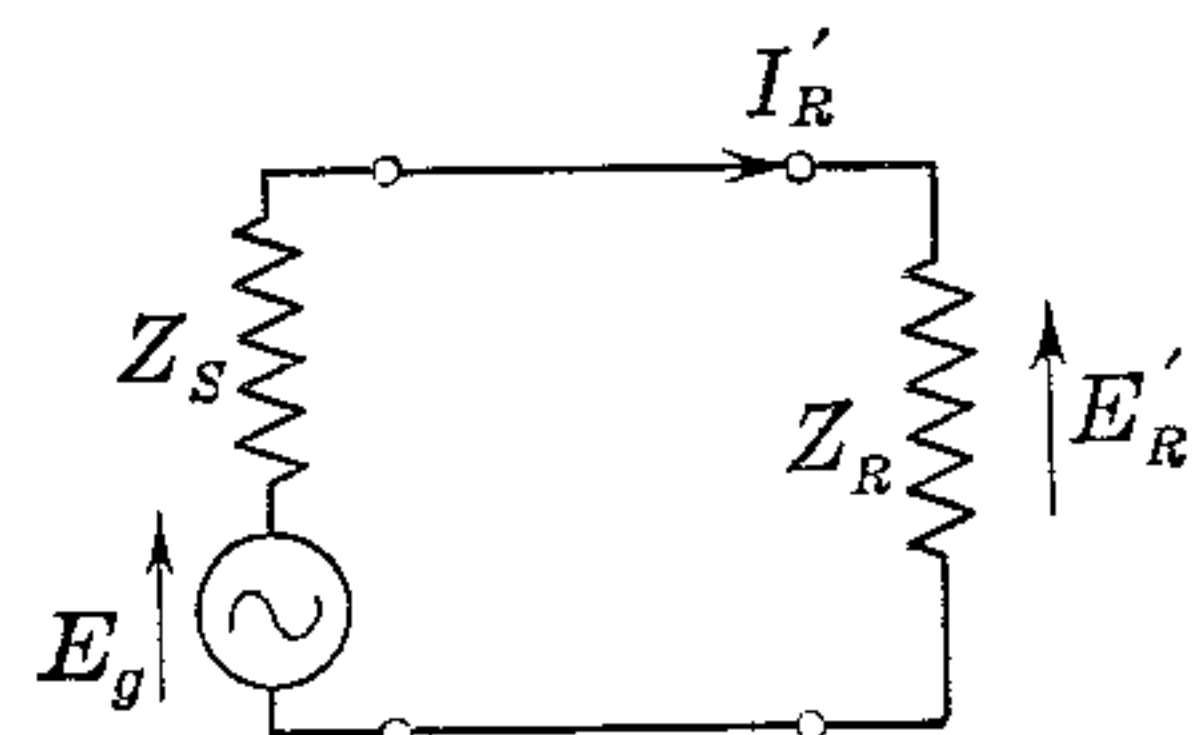


FIG. 12.—Source and load directly connected. Comparison with behavior for network of Fig. 4 places effect of line in evidence.

because it usually contributes the largest share to the magnitude of the voltage ratio. Its value is determined entirely by the propagation function. The last two factors indicate the effect upon the ratio which is due to the discrepancies between the terminal and the characteristic impedances. If either $Z_S = Z_0$ or $Z_R = Z_0$, these factors both become unity. Their effect is usually

small as compared to that of the exponential factor. Furthermore, the effect of the second-last factor is considerably greater than that of the last, particularly if the attenuation function α_1 is appreciable. This may be seen from the fact that the $r_S r_R$ -product in the last factor is multiplied by $e^{-2\alpha l}$, which is a very small number if the system is at all long or the attenuation relatively large. Thus the factors on the right-hand side of (128) are seen to be arranged in the order of their importance in affecting the net result. For a first approximation the first two will usually suffice. If this value is to be improved by partially taking the discrepancies between Z_S , Z_R , and Z_0 into account, the third factor may be included. The last factor then adds a final refinement to the numerical result. This is the situation in most practical cases.

Although the ratio (128) is useful in many ways, it really does not specifically indicate the effect which the transmission line itself has

upon the system as a whole. In order to see this effect, we must form the ratio of that voltage which would be received if Z_S and Z_R were directly connected, to the voltage which actually appears. Fig. 12 illustrates the direct-connected situation. Here the effect of the leads between Z_S and Z_R is assumed to be negligible. Current and voltage at the receiving end are denoted by I_R' and E_R' respectively. Under these conditions we obviously have

$$\frac{E_g}{E_R'} = \frac{Z_S + Z_R}{Z_R}, \quad (129)$$

which illustrates the significance of the first factor in (128). From these two relations it now follows that

$$\frac{E_R'}{E_R} = e^{\alpha l} \cdot (1 - r_S r_R)^{-1} \cdot (1 - r_S r_R e^{-2\alpha l}). \quad (130)$$

This ratio shows the effect of inserting the transmission line between the terminal impedances Z_S and Z_R and is, therefore, called the *insertion ratio*.

In a similar way we find from (61a) the following ratio of sending- to receiving-end current:

$$\frac{I_g}{I_R} = \frac{I_S}{I_R} = e^{\alpha l} \cdot (1 - r_R)^{-1} \cdot (1 - r_R e^{-2\alpha l}), \quad (131)$$

in which the order of importance of the factors again tapers from left to right.

For the ratio of generator voltage to received current, we have merely to recall that

$$I_R Z_R = E_R, \quad (132)$$

and thus

$$\frac{E_g}{I_R} = \frac{Z_R E_g}{E_R}, \quad (133)$$

so that this is immediately obtained from (128).

Since

$$\left. \begin{aligned} I_R' Z_R &= E_R' \\ I_R Z_R &= E_R \end{aligned} \right\}, \quad (134)$$

and

we see that

$$\frac{I_R'}{I_R} = \frac{E_R'}{E_R}. \quad (135)$$

Hence the insertion ratio for the current is the same as for the voltage.

In comparing the input-to-output voltage ratio (128) with (131) for the current, the reader will note a lack of similarity in these forms.

ATHERFORCE

This is due to the fact that although the generator current equals the sending-end current, the same is not true for the voltage. In other words, the forms (128) and (131) are not on a comparable basis. The voltage ratio comparable to (131) is that for the sending- to receiving-end voltage. This may be found either directly from (60a) or more easily by setting $Z_S = 0$ in (128). The reader should convince himself of this by carrying out the former and comparing with the result of the latter, which is

$$\frac{E_S}{E_R} = e^{\alpha l} \cdot (1 + r_R)^{-1} \cdot (1 + r_R e^{-2\alpha l}). \quad (128a)$$

This is the voltage ratio which is on the same basis with (131). We see that neither depends upon Z_S .

In section 6 we saw that, whenever the line is terminated in its characteristic impedance, the voltage and current maintain the same phase relationship at all points in the system. In practice this condition is usually fairly nearly realized, so that the magnitude of the volt-ampere ratio then represents the corresponding power ratio.

For sending to receiving end, the volt-ampere ratio is given by the product of (131) and (128a), thus

$$\frac{E_S I_S}{E_R I_R} = e^{2\alpha l} \cdot (1 - r_R^2)^{-1} \cdot (1 - r_R^2 e^{-4\alpha l}). \quad (136)$$

For the volt-ampere insertion ratio we have by (130) and (135)

$$\frac{E_R' I_R'}{E_R I_R} = e^{2\alpha l} \cdot (1 - r_S r_R)^{-2} \cdot (1 - r_S r_R e^{-2\alpha l})^2. \quad (137)$$

It should be remembered that the magnitudes of these expressions become power ratios only when the line is terminated in Z_0 , in which case we obviously get:

$$\frac{E_S}{E_R} = \frac{I_S}{I_R} = \sqrt{\frac{E_S I_S}{E_R I_R}} = \sqrt{\frac{E_R' I_R'}{E_R I_R}} = e^{\alpha l}. \quad (138)$$

In practice it has been found that it is more convenient to work with the logarithms of the above ratios than with the numerical ratios themselves. There are several reasons for this. We saw that in general the voltage, current, and volt-ampere ratios consist of products of various factors. The logarithm of such a product may be written as the sum of logarithms of the individual factors involved. This form evidently separates the contributions of the various factors in an additive manner and thus emphasizes their individual effects more strikingly.

For example, taking the natural logarithm of the voltage insertion ratio (130) we have

$$\ln \left(\frac{E_R'}{E_R} \right) = \alpha l - \ln (1 - r_S r_R) + \ln (1 - r_S r_R e^{-2\alpha l}). \quad (130a)$$

Here the magnitudes of the successive terms usually taper from left to right in the manner already pointed out. Since the ratio itself is in general complex, the terms on the right of (130a) are also complex. The sum of the real parts indicates the net attenuation; the sum of the imaginary parts equals the net phase displacement. Separation of the first term in this way is very simple, and often gives the major contribution to both attenuation and phase shift, namely

$$\ln \left(\frac{E_R'}{E_R} \right) = \alpha_1 l + j\alpha_2 l \quad (\text{approximately}). \quad (130b)$$

If we are interested only in the attenuation, as we frequently are, then we need consider only the logarithm of the absolute value of the ratio. Indicating absolute values by enclosing the corresponding complex quantities between vertical lines, we have

$$\ln \left| \frac{E_R'}{E_R} \right| = \alpha_1 l - \ln |1 - r_S r_R| + \ln |1 - r_S r_R e^{-2\alpha l}|, \quad (130c)$$

which is spoken of as the *insertion loss*.

Similar expressions apply to the other ratios discussed above. The absolute values of these ratios are called **loss ratios**, and their logarithms are simply spoken of as **losses**. The separate terms, such as those on the right-hand side of (130c), are sometimes given separate names. In fact, the separation into terms is frequently carried still farther than has been done here, but there is nothing in this process which is of fundamental interest. The term $\alpha_1 l$ may be called the **attenuation loss** since it depends upon the attenuation function of the line alone. The rest are due to reflections at the line ends, and the process of designating them by names¹ is purely a matter of personal taste which may vary with the individual and hence is difficult to standardize.

If the line is dissipationless and properly terminated, the loss ratio is unity and its logarithm zero. Thus we say that the loss is zero. This again illustrates the advantage in working with the logarithmic ratios instead of the ratios themselves. Zero loss means a ratio whose absolute value is unity. A positive loss represents a loss ratio greater than unity, i.e., a decrease in the magnitude of the quantity involved.

¹ A fairly common usage is to call the second term in (130c), inclusive of its minus sign, the *reflection loss*, and the third term the *interaction loss* since it involves not only the reflection coefficients but also the propagation function.

If the loss ratio is less than unity, as it may be under certain conditions pointed out earlier, then the logarithm is negative. The loss is then said to be *negative*, or is also spoken of as a *gain*. The extremes of infinite and zero loss ratio become, in the logarithmic sense, infinite loss and infinite gain, respectively. The logarithmic function is thus seen to be more symmetrical with respect to loss and gain than the ratio itself.

There is still another reason why the logarithmic mode of expression is preferable. This has to do with the observed fact that the sensitivity of the human ear approximately obeys a square law. That is, a given sound will seem twice as loud when the pressure variation has reached the square of its original value. This is, of course, the same as to say that the logarithm of the original pressure variation has doubled. Thus the logarithmic ratio is a more natural way of expressing the intensity level of a received signal.

18. Attenuation units. In practical work it is convenient to have a unit in terms of which a given logarithmic ratio may be expressed, and to designate such a unit by an appropriate name. When the natural or Napierian system of logarithms is used (as has been done in all the above work) the logarithmic ratio is said to be expressed in **napiers**.¹ This name for the natural loss unit is chosen in honor of the inventor of the natural system of logarithms. Thus if the natural logarithm of a ratio equals, say, 2, we say that the loss for this ratio is two napiers. It happens, however, that the size of this natural unit is inconveniently large. In telephone work it has been found more convenient to express the loss in terms of the following definition involving the Briggs or common system of logarithms

$$\text{Loss} = 20 \log \left| \frac{E_1}{E_2} \right| = 20 \log \left| \frac{I_1}{I_2} \right|, \quad (139)$$

where "log" is written for the common system to distinguish it from "ln" which refers to the natural, and the subscripts 1 and 2 on E and I are used to indicate these ratios in more general terms. The factor 20 is arbitrarily inserted in order to give the unit a desired size. Historically the size of the unit was determined by the attenuation inherent in a mile of what in the early days was known as "standard cable." In that day the theory of transmission lines had not yet been developed to the point where engineers could make effective use of it, and the idea of logarithmic ratios was not available. When the theory became more thoroughly established, however, and a more useful definition for attenuation or loss was decided upon, the form (139) with the

¹ This is also written *neper*, which is said to be the original spelling of the name

factor 20 was adopted because it led to a unit whose size most nearly approximated that for the old definition. The size of the latter was, of course, subject to a variation with frequency, but in the earlier days telephone lines were used only for voice frequencies, which were considered to be sufficiently well represented by an average value of 1,000 cycles per second. The loss in a mile of "standard cable" was understood to be referred to that frequency. The definition (139) is free from a frequency variation, and the unit size is still comparable to the old, so that engineers who are accustomed to thinking in terms of the old magnitude need not be inconvenienced by the change.

The original name for the unit defined by (139) was the **transmission unit**, which was abbreviated by the letters **T. U.** For properly terminated systems the relation (138) holds, and the magnitudes of the volt-ampere ratios become power ratios. We may then amplify the definition (139) and write

$$\begin{aligned} \text{Number of T.U.} &= 20 \log \left| \frac{E_1}{E_2} \right| = 20 \log \left| \frac{I_1}{I_2} \right|, \\ &= 10 \log \left| \frac{E_1 I_1}{E_2 I_2} \right| = 10 \log \frac{P_1}{P_2} \end{aligned} \quad (139a)$$

where P_1 and P_2 are used to designate power.

More recently another unit was introduced and named in honor of Alexander Graham Bell. This unit was called the **bel** and defined by

$$\text{Number of bels} = \log \frac{P_1}{P_2}. \quad (140)$$

This definition, which leaves off the factor 10, obviously leads to a unit size which is ten times as large as the T.U. Since this size is inconvenient, it was decided to divide it into tenths and express loss in terms of tenths of bels or **decibels**. These, however, are again coincident in size with the transmission units; i.e.,

$$\text{Number of decibels} = \text{Number of T.U.}$$

The term decibel is abbreviated as **db**, and the final upshot of the matter is that db and T.U. amount to the same thing.

In order to determine the relation between napiers and decibels, we merely have to compare their respective definitions; for example

$$\text{Number of napiers} = \ln \left| \frac{E_1}{E_2} \right|,$$

while

$$\text{Number of db or T.U.} = 20 \log \left| \frac{E_1}{E_2} \right|.$$

ETHERFORCE

If we recall that

$$\log x = \ln x \cdot \log e = 0.4343 \ln x,$$

then we have

$$\begin{aligned} \text{Number of db or T.U.} &= 20 \times 0.4343 \ln \left| \frac{E_1}{E_2} \right| \\ &= 8.686 \cdot \text{Number of napiers.} \end{aligned} \quad (141)$$

Notice that the words "number of" are important. If we replace these by the words "size of," then the factor 8.686 appears on the other side of the equation, thus

$$8.686 \cdot \text{Size of a db or T.U.} = \text{Size of a napier.} \quad (141a)$$

In using such conversion equations the student should always be sure whether the words "number of" or "size of" are implied in the particular case in hand.

PROBLEMS TO CHAPTER II

2-1. An open telephone line has the parameter values: $R = 10.4$; $L = 0.00367$; $C = 0.00835 \cdot 10^{-6}$; $G = 0.8 \cdot 10^{-6}$. At the angular frequencies $\omega = 1000$; $\omega = 5000$; $\omega = 10,000$; $\omega = 20,000$, calculate the following:

- The attenuation function α_1 .
- The phase function α_2 .
- The wave length λ .
- The phase velocity v .
- The characteristic impedance Z_0 in magnitude and angle. Make sketches of these quantities versus frequency.

2-2. Repeat the calculations and plots indicated in Problem 2-1 for the cable having the following parameter values: $R = 83.2$; $L = 0.001$; $C = 0.062 \cdot 10^{-6}$; $G = 0.868 \cdot 10^{-6}$.

2-3. In Problems 2-1 and 2-2, neglect dissipation, and calculate and plot the indicated quantities. Compare with the corresponding dissipative values.

2-4. The open line of Problem 2-1 is considered at a frequency of 796 cycle per second and at this frequency is one wave length long. Determine the polar plots for the voltage and current for the following terminal conditions:

- $Z_S = Z_R = 700$.
- $Z_S = 700$; $Z_R = 2000$.
- $Z_S = 0$; $Z_R = \infty$.
- $Z_S = 0$; $Z_R = j 1000$.

2-5. Repeat Problem 2-4 with the cable data of Problem 2-2 for the terminal conditions:

- $Z_S = Z_R = 500$.
- $Z_S = 500$; $Z_R = 2000$.
- $Z_S = 0$; $Z_R = 0$.
- $Z_S = 0$; $Z_R = j 500$.

2-6. For the open line of Problem 2-1 neglect R and G and determine the ellipse diagrams for voltage and current for the terminal conditions specified in Problem 2-4. Compare with the polar plots obtained for that problem.

2-7. The open line of Problem 2-1 is 200 miles long and is terminated at either end in resistances of 700 ohms. Calculate the logarithmic insertion ratio (130a) for the angular frequencies $\omega = 1000$ to $\omega = 20,000$ at intervals sufficiently close to plot curves of the real and imaginary parts of each of the three terms in this expression as well as of their respective sums.

2-8. Repeat Problem 2-7 for the cable whose data are specified in Problem 2-2 for a length of 50 miles and terminal resistances of 500 ohms each.

2-9. A pair of No. 10 copper conductors spaced 8 in. on centers is excited from a source having a frequency corresponding to a wave length of 30 meters. For the terminal conditions $Z_S = 0$, $Z_R = \infty$; $Z_S = 0$, $Z_R = 0$, determine the standing wave patterns for voltage and current for lengths corresponding to the following fractions of a wave length: $\frac{3}{4}$, $\frac{7}{8}$, 1. Neglect dissipation. Repeat for the line one wave length long but terminated at the far end in a pure inductance of one microhenry. Determine approximately the effect of line dissipation on these results.

2-10. Consider the line of Problem 2-9 one wave length long with $Z_S = 0$; $Z_R = 2000$, 1000, 500 ohms pure resistance. For each case determine the ellipse diagrams for voltage and current. Neglect dissipation of the line.

2-11. The line of Problem 1-1 is directly connected to a source of negligible internal impedance, and is terminated at the far end in a pure inductance. If the line is one wave length long, how large should the terminal inductance be so that the maxima of the voltage standing waves will not exceed $\sqrt{2}$ times the maximum voltage of the source? If the terminal inductance is made larger than this value, will the maxima become larger or smaller?

2-12. In a Lecher wire system which is used for frequency measurement the high-impedance-indicating device is connected to the far end. A piece of copper wire is bridged across the line and moved gradually from the far end toward the source. When a distance of one-quarter wave length from the far end is reached, the indicator suddenly registers. Discuss the behavior of the system under this critical condition so as to show why it functions in this manner.

2-13. Utilizing the relations (114a), show that the impedance looking into a line terminated in Z_R may be expressed as

$$Z(0) = \frac{\alpha Z_R + \beta}{\epsilon Z_R + \mathcal{D}}.$$

Making use of the results of Problem 2-2, calculate the impedance looking into this cable at the frequencies indicated and for all combinations of the following lengths and terminal impedance values:

- $$\begin{aligned} Z_R &= 500 \text{ ohms pure resistance.} \\ Z_R &= 2000 \text{ ohms pure resistance.} \\ l &= 50 \text{ miles.} \\ l &= 100 \text{ miles.} \end{aligned}$$

2-14. A 100-mile length of open line as specified in Problem 2-1 is joined directly to 50 miles of cable as specified in Problem 2-2, and the latter is terminated in a resistance of 500 ohms. At the frequencies indicated in 2-1, calculate:

- The reflection coefficient at the end of the cable.
- The reflection coefficient at the end of the line.
- The ratio of sending-end voltage to that at the point where the line joins the cable.
- The overall voltage ratio.

CHAPTER III

PROPAGATION AND CHARACTERISTIC IMPEDANCE
FUNCTIONS OF THE LONG LINE

1. Ideal behavior. From the discussion in the preceding chapter it should be clear that the quality of the transmission line as a communication facility depends entirely upon the functions α and Z_0 defined by equations (43a) and (53) respectively. According to these equations we see that the behavior of α and Z_0 versus frequency depends upon all four line parameters R , L , G , and C . It is important to study this dependence upon the line parameters so that we may be able to determine the most favorable combination of values for them, as well as to establish criteria for the readjustment of their values in order that the line behavior in a given case may meet a prescribed degree of quality. Before going into the details of this study, it is well to review briefly again the most desirable forms for these functions so that we may have them before us as a comparative standard.

In section 6 of the preceding chapter we saw that, whenever the line is terminated in its characteristic impedance, reflections are absent. This is a most desirable condition because reflections not only cause echoes, which are a particular form of combined phase and amplitude distortion,¹ but also give rise to such distortion even though the propagation function meets all the conditions for perfect transmission. Furthermore, it was pointed out in the first chapter that radiation in the steady state was due to, and took place in the vicinity of, the line ends, but that this phenomenon was wholly absent in the infinitely long line. When a line is terminated in its characteristic impedance, it is virtually infinitely long, so that the performance proceeds as though the termination did not exist. The greater the mismatch at a terminal, the greater will be the chance for the induction of electromotive forces in neighboring circuits. This may become noticeable at higher frequencies such as are used in carrier telephony. Transpositions are not effective in counteracting this phenomenon. Lastly, proper termination is desirable from the standpoint of transmission efficiency since under these conditions all the transmitted energy is accepted by the terminal device.

¹ This point will be discussed in Chapter XI.

From the practical standpoint, the condition of proper termination is most easily met when the characteristic impedance is a real constant. This will, incidentally, also make the transmitted energy a maximum, as may be seen independently by considering a source with an internal impedance Z_i feeding a load impedance Z_x . Assuming the voltage of the source unity, and writing

$$\left. \begin{aligned} Z_i &= R_i + jX_i \\ Z_x &= R_x + jX_x \end{aligned} \right\},$$

the power expended in Z_x becomes

$$P = \frac{R_x}{(R_i + R_x)^2 + (X_i + X_x)^2}.$$

Differentiating partially with respect to X_x and setting this equal to zero gives

$$X_x = -X_i.$$

Substituting this into the expression for P , the subsequent condition for maximum power with respect to R_x gives

$$R_x = R_i.$$

Thus if the two impedances are pure resistances, the condition for maximum power transfer is simply that they shall be equal. When they have reactive components, however, then the latter should be equal and opposite. This is equivalent to a resonance condition. It shows that when Z_i and Z_x are complex and equal, the power transfer is not a maximum. In the transmission line, Z_0 is the one impedance and the termination the other. The condition for no reflection can be realized simultaneously with that for maximum power transfer only when Z_0 reduces to a constant. Hence this is the ideal for the characteristic impedance function.

From (71) and (72) we see that if the line is terminated in Z_0 at both ends we have

$$E(x) = \frac{E_g}{2} \cdot e^{-\alpha x}, \quad (71a)$$

$$I(x) = \frac{E_g}{2Z_0} \cdot e^{-\alpha x}. \quad (72a)$$

Assuming a constant Z_0 , the behavior is then entirely governed by α . When the steady state consists of a linear superposition of a number of harmonic components in the Fourier sense, then the distortionless transmission of the resulting periodic function obviously requires that the relative attenuation and the phase velocities for the individual

components all be alike. The requirement for no amplitude distortion will evidently be met if the real part of α is a constant, while the phase velocity according to (80a) will be independent of frequency if the imaginary part of α is a linear function of frequency. This defines the ideal behavior for the attenuation and phase functions.

The reader should note well that the conditions for a distortionless transmission system involve *both* α and Z_0 . So long as the latter is not constant, the system will possess some amplitude and phase distortion even though α_1 be constant and α_2 a linear function of frequency. On the other hand, if the variation in Z_0 with respect to Z_S and Z_R is not very great, i.e., if the terminal conditions are reasonably good, then in any practical system the behavior is almost wholly governed by α . If the error involved in this assumption is of doubtful magnitude, the reflection effects must, of course, be calculated according to the methods already discussed.

So far the only distortionless case that we have said anything about is that for neglected dissipation. There $Z_0 = \sqrt{L/C}$, $\alpha_1 = 0$, and $\alpha_2 = \omega\sqrt{LC}$. In general, Z_0 , α_1 , and α_2 are rather complicated functions of frequency and in some cases differ very materially from their ideal behavior. With suitable readjustments in the line parameters, however, they may in general be made to assume their ideal characteristics to within a reasonable degree of error over any prescribed finite region of frequencies. Since for any one channel of communication only a finite frequency range is required, it is necessary only that the functions Z_0 , α_1 , and α_2 conform sufficiently well to their ideal standards within this range. If the neighboring regions are not utilized, it is needless to worry about the behavior of these functions there. In practice it is usually much simpler to readjust matters so that the behavior will be good over a narrow frequency range than it is to do this for a wide one. The location of the range in the frequency spectrum also has much to do with the practical difficulties involved. Whatever the case may be, the problem in general is to meet certain prescribed tolerances regarding amplitude and phase distortion over a given range of frequencies, and to do this with the minimum readjustments in the line parameters. To do more than is required would be a waste. The problem is to determine that set of readjustments which will just meet the requirements. This, at least, is the logical statement of the problem so far as the line itself is concerned. In present-day practice it is found more economical to correct for a large share of the shortcomings of the line by means of special networks which are inserted at the terminals. An appreciation of the economies involved in such practice requires first that the prob-

lem without auxiliary networks be understood. This we shall now undertake to present.

2. Separation of the propagation function. We shall begin with the discussion of the propagation function. From (43a) we have

$$\alpha = \sqrt{(R + jL\omega)(G + jC\omega)} \left. \vphantom{\alpha} \right\} \quad (142)$$

$$= \alpha_1 + j\alpha_2$$

Obviously the first problem is to separate the radical into its real and imaginary components. Squaring both sides of (142) we get

$$\alpha_1^2 - \alpha_2^2 + 2j\alpha_1\alpha_2 = (RG - LC\omega^2) + j\omega(LG + RC). \quad (143)$$

Equating reals and imaginaries, this gives

$$\left. \begin{aligned} \alpha_1^2 - \alpha_2^2 &= RG - LC\omega^2 \\ 2\alpha_1\alpha_2 &= (LG + RC)\omega \end{aligned} \right\} \quad (144)$$

If we square each of these, add the results, and after some factoring again extract the square root, we find

$$\alpha_1^2 + \alpha_2^2 = \sqrt{(R^2 + L^2\omega^2)(G^2 + C^2\omega^2)}. \quad (145)$$

With this and the first equation (144) we can easily obtain α_1^2 and α_2^2 by addition and subtraction respectively. Thus we get

$$\alpha_1^2 = \frac{1}{2} [(RG - LC\omega^2) + \sqrt{(R^2 + L^2\omega^2)(G^2 + C^2\omega^2)}], \quad (146)$$

and

$$\alpha_2^2 = \frac{1}{2} [(LC\omega^2 - RG) + \sqrt{(R^2 + L^2\omega^2)(G^2 + C^2\omega^2)}]. \quad (147)$$

The square roots of these expressions formally at least represent the desired separation. It may be easily appreciated that the resulting forms are rather awkward from the standpoint of their discussion. For a given set of numerical values of R , L , G , and C , the process of plotting α_1 and α_2 versus frequency is, of course, a straightforward matter although somewhat long and tedious. What we wish to accomplish, however, is to discuss these functions analytically so that the relative effects of the various parameters will become more evident. In order to carry this out, various changes of variable as well as the introduction of several new parameters will be made. We hope that this process will not discourage the reader from continuing.

3. Discussion of the attenuation and phase functions. The primary object in the following manipulations is to put the analytic expressions for α_1 and α_2 in such a form that the departure from their ideal behavior in the general case becomes most evident. In attempting to carry out this idea we shall express these functions as products of their ideal values, and such factors which in the ideal case become unity.

Several substitutions are convenient in this connection. These are

$$\left. \begin{aligned} \frac{R}{L} &= a \\ \frac{G}{C} &= b \\ x &= \frac{\omega}{\sqrt{ab}} \end{aligned} \right\}. \quad (148)$$

Then (146) and (147) may be written

$$\alpha_1^2 = \frac{RG}{2} \left\{ (1 - x^2) + \sqrt{\left(\frac{a}{b} + x^2\right)\left(\frac{b}{a} + x^2\right)} \right\} \quad (146a)$$

$$\alpha_2^2 = \frac{LC\omega^2}{2} \left\{ (1 - x^{-2}) + \sqrt{\left(\frac{a}{b} + x^{-2}\right)\left(\frac{b}{a} + x^{-2}\right)} \right\}. \quad (147a)$$

The factors RG and $LC\omega^2$ are the ideal values referred to. The remaining factors indicate the departure from the ideal. The discussion which follows concerns itself chiefly with these remaining factors. Here the additional substitutions are useful

$$\left. \begin{aligned} m &= \frac{1}{2} \left(\sqrt{\frac{a}{b}} + \sqrt{\frac{b}{a}} \right) \\ n &= \frac{1}{2} \left(\sqrt{\frac{a}{b}} - \sqrt{\frac{b}{a}} \right) \end{aligned} \right\}. \quad (149)$$

With these we find that (146a) and (147a) become

$$\alpha_1 = \sqrt{RG} \cdot f_1(x) \quad (146b)$$

$$\alpha_2 = \omega\sqrt{LC} \cdot f_2(x) \quad (147b)$$

with

$$f_1(x) = \frac{1}{\sqrt{2}} \left\{ 1 - x^2 + \sqrt{1 + 2(m^2 + n^2)x^2 + x^4} \right\}^{1/2} \quad (150)$$

$$f_2(x) = \frac{1}{\sqrt{2}} \left\{ 1 - x^{-2} + \sqrt{1 + 2(m^2 + n^2)x^{-2} + x^{-4}} \right\}^{1/2}. \quad (151)$$

The functions $f_1(x)$ and $f_2(x)$ contain all that is pertinent to the variation of α_1 and α_2 . They have several rather interesting properties. One of these is evident from (150) and (151), namely that

$$f_2(x) = f_1\left(\frac{1}{x}\right), \quad (152)$$

which states that the branch of $f_1(x)$ for the region $0 \leq x \leq 1$ is reciprocal to that of $f_2(x)$ for $1 \leq x \leq \infty$, and vice versa. Hence, if $f_1(x)$ is known for $0 \leq x \leq \infty$, then $f_2(x)$ may be found by reciprocation; or if $f_1(x)$ and

$f_2(x)$ are known in either of the regions $0 \leq x \leq 1$ or $1 \leq x \leq \infty$, then the balance of these functions may be got by geometrical construction.

The relationship between $f_1(x)$ and $f_2(x)$ goes even farther than this, as may be more readily seen if we make another change of variable, namely let

$$\left. \begin{aligned} y &= \frac{1}{2} \left(x - \frac{1}{x} \right) \\ x &= \sqrt{y^2 + 1} + y \end{aligned} \right\}. \quad (153)$$

Then we have

$$f_1(x) = x^{1/2} (\sqrt{m^2 + y^2} - y)^{1/2}, \quad (150a)$$

$$f_2(x) = x^{-1/2} (\sqrt{m^2 + y^2} + y)^{1/2}. \quad (151a)$$

From these we see that

$$f_1(x) \cdot f_2(x) = m, \quad (154)$$

which shows that $f_2(x)$ is the reciprocal of $f_1(x)$ with respect to the parameter m . Hence the branches of $f_1(x)$ and $f_2(x)$ are not only reciprocal about $x = 1$ for the regions $0 \leq x \leq 1$ and $1 \leq x \leq \infty$, but are also reciprocal about $f_1(1) = f_2(1) = \sqrt{m}$ within each region. This means that, if, for example, $f_1(x)$ is determined for $0 \leq x \leq 1$, then the balance of $f_1(x)$ and the whole of $f_2(x)$ can be found by geometrical construction alone!

Another aid toward plotting $f_1(x)$ and $f_2(x)$ is found if we differentiate (154) with respect to x . This gives

$$f_1(x) \cdot f_2'(x) + f_2(x) \cdot f_1'(x) = 0, \quad (155)$$

where primes denote differentiation with respect to the argument. If we note that

$$f_1(1) = f_2(1) = \sqrt{m} \quad (156)$$

we have for the point $x = 1$

$$f_2'(1) = -f_1'(1), \quad (155a)$$

i.e., the functions have equal and opposite slopes at this point. For $f_1(x)$ this slope is found most easily by considering twice the square of (150). Differentiating this with respect to x , we find

$$4f_1 \cdot \frac{df_1}{dx} = -2x + \frac{1}{2}(1 + 2(m^2 + n^2)x^2 + x^4)^{-1/2} \cdot (4x(m^2 + n^2) + 4x^3),$$

and setting $x = 1$, we have with the help of (156)

$$f_1'(1) = \frac{1}{2}(m^{1/2} - m^{-1/2}). \quad (157)$$

A further aid in plotting is found in series expansions of $f_1(x)$ and $f_2(x)$ about the points $x = 0$ and $x = \infty$. These will show how the

functions behave in the vicinity of the origin and infinity. Just a few terms will suffice to accomplish this. We find from (150) and (152) or (154)

$$\left. \begin{aligned} \text{About } x = 0: f_1(x) &= 1 + \frac{n^2 x^2}{2} + \dots \\ f_2(x) &= m \left(1 - \frac{n^2 x^2}{2} + \dots \right) \end{aligned} \right\} \quad (158)$$

and

$$\left. \begin{aligned} \text{About } x = \infty: f_1(x) &= m \left(1 - \frac{n^2}{2x^2} + \dots \right) \\ f_2(x) &= 1 + \frac{n^2}{2x^2} + \dots \end{aligned} \right\} \quad (159)$$

These show that the functions vary between the values 1 and m , and that they have zero slopes at the points $x = 0$ and $x = \infty$.

With the above information regarding $f_1(x)$ and $f_2(x)$, it is now a simple matter to plot these functions. This is done in Fig. 13, for which a

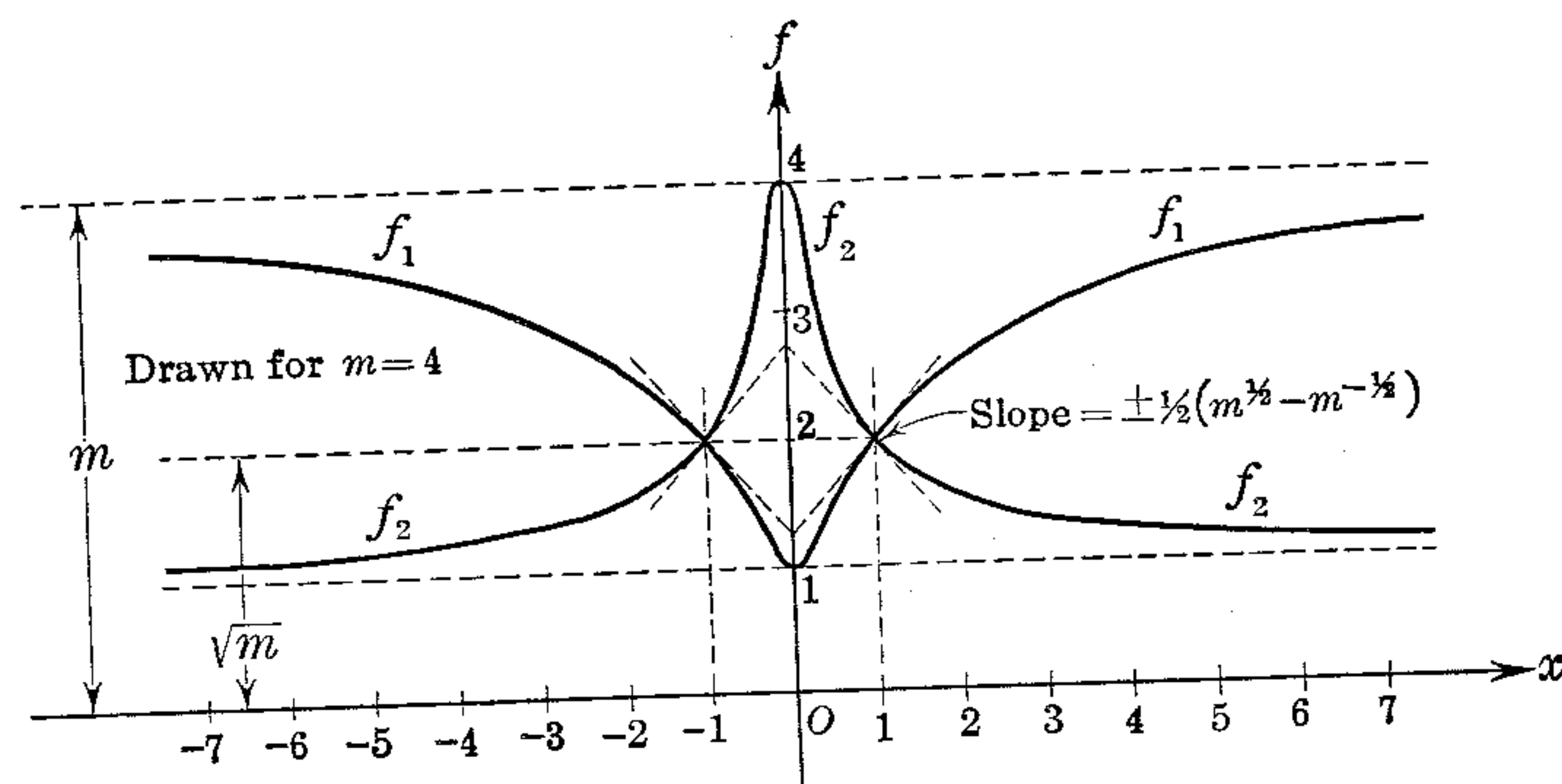


FIG. 13.—Functions showing departure from ideal behavior on the part of the attenuation and phase functions according to (146b) and (147b).

value $m = 4$ is arbitrarily chosen. The functions are plotted for negative as well as positive values of x in order to emphasize the fact that they are even functions. By (146b) and (147b) we, therefore, see that α_1 is an *even*, and α_2 an *odd*, function of frequency. Although negative frequencies have no physical significance, their consideration will be convenient later on in the transient analysis by means of the Fourier integral methods.

Finally in Fig. 14 we show plots of α_1 and α_2 corresponding to the plots of $f_1(x)$ and $f_2(x)$ in Fig. 13.

4. Amplitude and phase distortion. Assuming for the moment that the line is properly terminated, the functions α_1 and α_2 completely

control the amplitude and phase distortion for the system, while the functions $f_1(x)$ and $f_2(x)$ most conveniently indicate the magnitude of the distortion. If these functions were constants, the transmission would be distortionless. The amount of distortion in any case is best

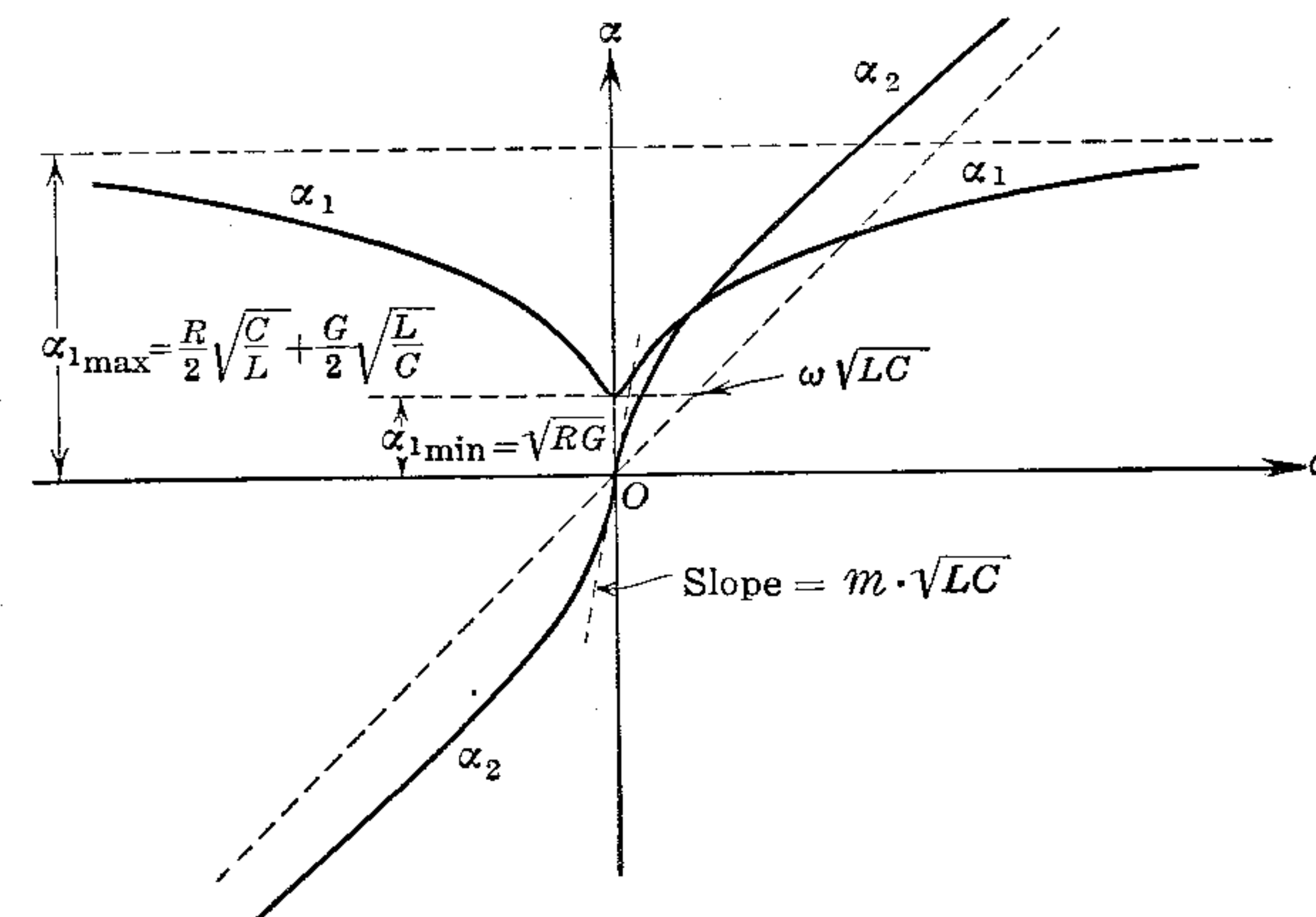


FIG. 14.—Plots of the attenuation and phase functions corresponding to the departures shown in Fig. 13.

expressed in terms of the percentage deviation of these functions from constant values over the essential frequency range for which the facility is to be used.

Fig. 13 shows that $f_1(x)$ has for its asymptote the value m , while $f_2(x)$ approaches unity with increasing frequency. It is also apparent

that most of the distortion occurs at the lower frequencies; for frequencies above that corresponding to about $x = 4$ or 5 the distortion is certainly not appreciable. The maximum amount of distortion in any case evidently depends upon the value of the parameter m .

If we examine its analytic

form as given by (149) we see that its value depends upon the ratio of a to b , which in turn are given by the line parameters, as expressed by (148). More precisely, m is equal to half the sum of $\sqrt{a/b}$ plus its reciprocal. Fig. 15 shows both the parameters m and n as functions of

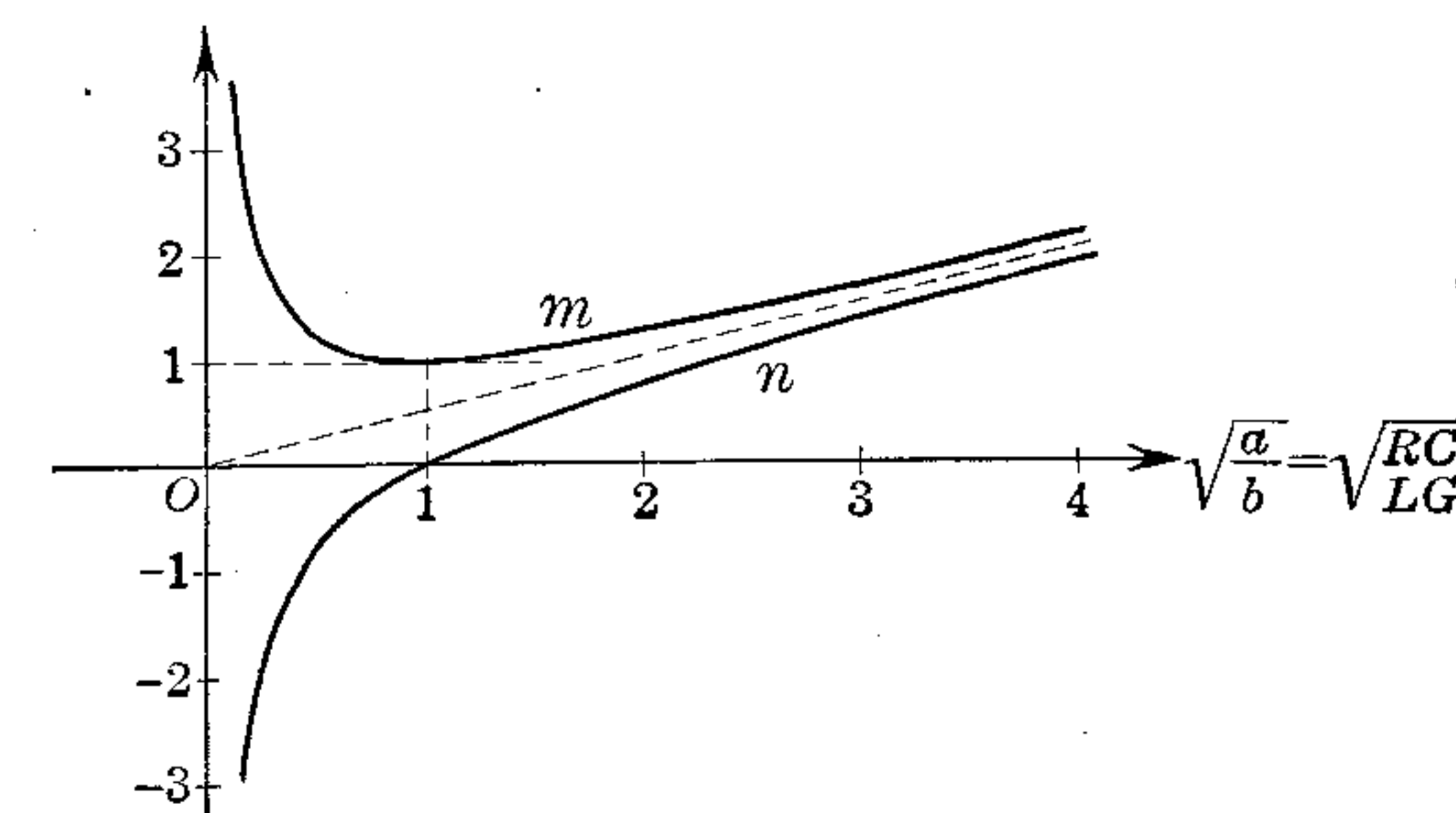


FIG. 15.—Dependence of the parameters (149) upon the line constants.

this square root. We see that m has a minimum value of unity which occurs for

$$\sqrt{\frac{a}{b}} = 1$$

or

$$a = b,$$

which by (148) is equivalent to

$$\frac{R}{L} = \frac{G}{C}. \quad (160)$$

For all other values of $\sqrt{a/b}$ it is larger than unity. Referring to Fig. 13 again we see that if $m = 1$ the upper asymptote coincides with the lower and both f_1 and f_2 reduce to unity at all frequencies! Thus, when (160) holds, the attenuation and phase functions follow their ideal behavior. We shall see in a later section of this chapter that the condition (160) also reduces Z_0 to a real constant. This then is the condition for which a dissipative line will be distortionless. We saw earlier that the non-dissipative line was distortionless. Now we see that the dissipative line may also be distortionless, but only if the line parameters satisfy the condition (160).

In practice this condition never exists. The usual situation is that

$$\frac{R}{L} > \frac{G}{C}. \quad (161)$$

Theoretically there are obviously four ways in which the condition (161) may be made to approach that given by (160). We might try to decrease R , increase L , increase G , or decrease C . Practically there is no way of decreasing the capacitance, and so this method is out of the question. G might be increased by using a poorer insulator, but although the functions $f_1(x)$ and $f_2(x)$ could be made unity by a sufficient increase in G , equation (146b) shows that the resulting attenuation would thereby be increased. Hence this process cannot be recommended. Increasing L is a good remedy but rather difficult to carry out in practice owing to the fact that the increase in inductance must be uniformly distributed. One way in which this is accomplished is by uniformly wrapping the conductors with a ribbon made of some material having a high magnetic permeability. This is done in some ocean cables, and is referred to as **uniform loading**. As this is a very expensive process, however, it is not resorted to unless there is no other way of meeting the situation. On land lines the problem of increasing L is approximately met by inserting the added inductance in small concentrated amounts at regular intervals. In this way a uniform dis-

tribution may obviously be approximated, the approximation becoming better the smaller the intervals. This process is referred to as **lump loading**. Unless the intervals are small enough, however, this method may introduce serious distortion at the higher frequencies. Since the expense involved increases materially as the intervals are shortened, it is necessary to determine carefully the maximum interval that may be tolerated for a given case. This determination, which involves some fundamental principles that we have not yet discussed, will be taken up in a later chapter.

Finally, by decreasing R we not only approach the distortionless condition, but also decrease the final value of the attenuation function α_1 as is evident from (146b). This is, therefore, the most desirable way of improving the situation. Since it involves a considerable increase in the quantity of copper needed for a given facility, it is quite a costly process. In cables, the increase in wire size reduces the number of pairs that can be placed in a given space. In congested areas this space factor definitely limits the extent to which this method of improvement may be carried. On open wire lines, on the other hand, the present tendency is to use more copper and omit line loading as a means for improvement. Loading practice is being confined more and more to cable facilities where the wire size is limited.

It should also be pointed out in this connection that in practice, line loading or increasing the line copper is not done primarily in order to attain or approximate the distortionless condition. The chief object is to reduce the attenuation at higher frequencies by reducing the asymptotic value of α_1 . Noting that the asymptote of $f_1(x)$ is m , that for α_1 , is, by (146b)

$$\sqrt{RG} \cdot m,$$

or using (148) and (149) this is

$$(\alpha_1)_{\max} = \frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}}. \quad (162)$$

It is interesting to compare this maximum value of the attenuation function with its minimum value which occurs at zero frequency and obviously is

$$(\alpha_1)_{\min} = \sqrt{RG}. \quad (163)$$

A study of these maximum and minimum values as functions of the parameters R , L , and G brings out some useful facts regarding the relative merits of readjusting their magnitudes in attempting to approach the distortionless condition or reducing the average attenuation.

ETHERFORCE

This we can very effectively illustrate by considering the following data which are approximately representative of an open telephone line. Here

$$R = 10; L = 0.004; G = 0.8 \cdot 10^{-6}; C = 0.008 \cdot 10^{-6}$$

per mile. Using these values in (162) and (163), the curves of Figs. 16, 17, and 18 are obtained for $(\alpha_1)_{\max}$ and $(\alpha_1)_{\min}$ as functions of R , L , and G respectively. In these figures, dotted ordinates are drawn at the points corresponding to the original parameter values, and again at those values for which the distortionless line condition is satisfied. It will be seen that the two curves become tangent at these points.

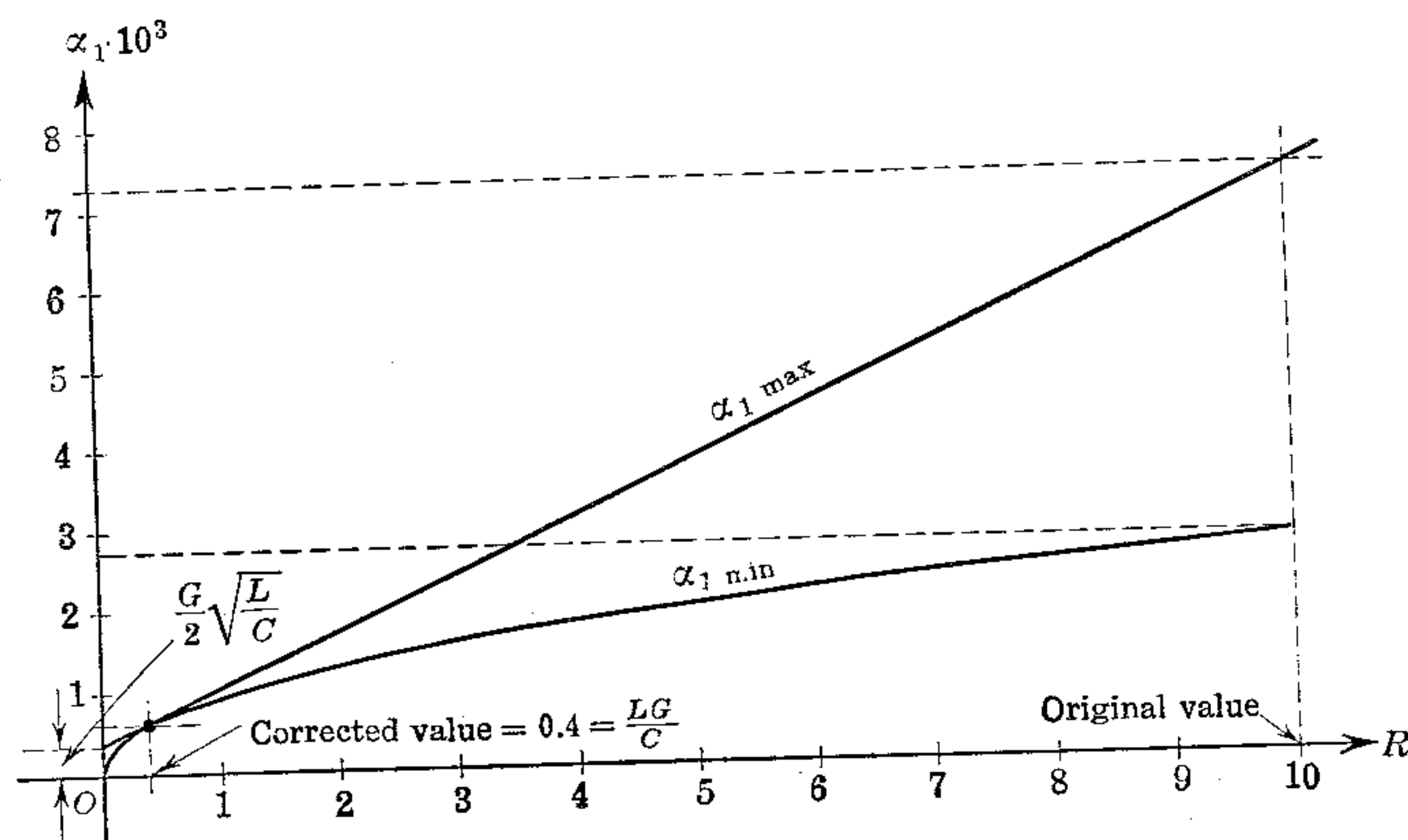


FIG. 16.—Dependence of the maximum and minimum values of attenuation upon the resistance parameter for constant values of L , G , and C .

The vertical separation at any point represents the maximum variation of the attenuation function with frequency for that parameter value.

Fig. 16 shows that decreasing R not only brings the two curves closer together, but also materially decreases the attenuation at both the low and high-frequency limits. From Fig. 17 we see that increasing the inductance decreases the high-frequency value but leaves the minimum value unchanged. This method of approaching the distortionless condition is, therefore, not quite as desirable as already pointed out, although the considerable decrease in $(\alpha_1)_{\max}$ which is accomplished by loading is extremely useful in supporting the higher frequencies. In fact, line loading in the early days was chiefly heralded for this very reason. Fig. 17 brings out another very interesting point. We see that increasing L from its initial value is most effective throughout the first

few hundred per cent, say up to the value 0.03. Beyond this, very little is to be gained by a further increase, either in the reduction of $(\alpha_1)_{\max}$ or in the attainment of the distortionless condition. Hence it would

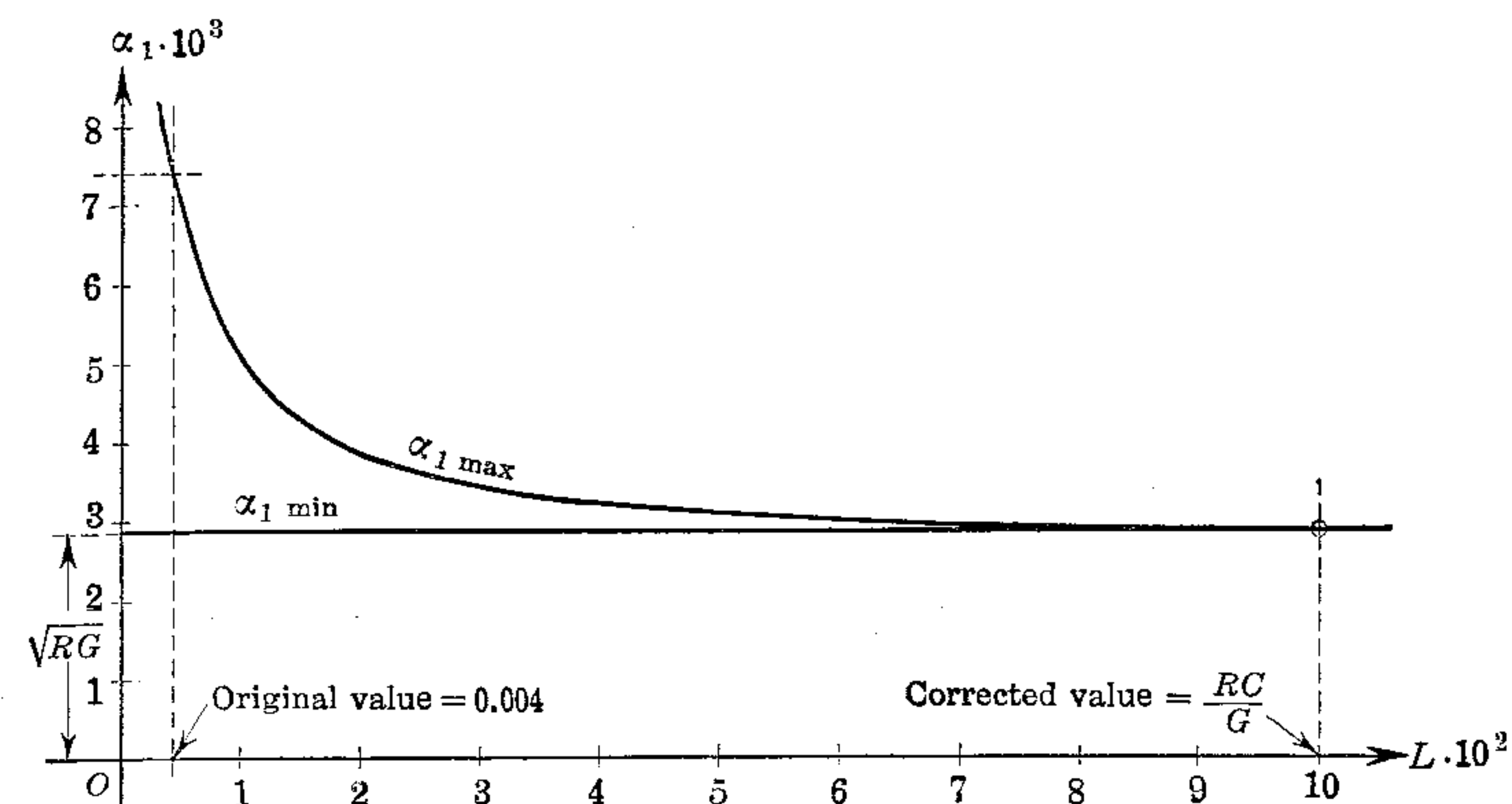


FIG. 17.—Dependence of the maximum and minimum values of attenuation upon the inductance parameter for constant values of R , G , and C .

hardly be economical to go beyond this value in any corrective procedure. Even a comparatively small amount of loading may thus do a very noticeable amount of good.

Fig. 18 shows how poor a policy it is to increase G in an attempt to

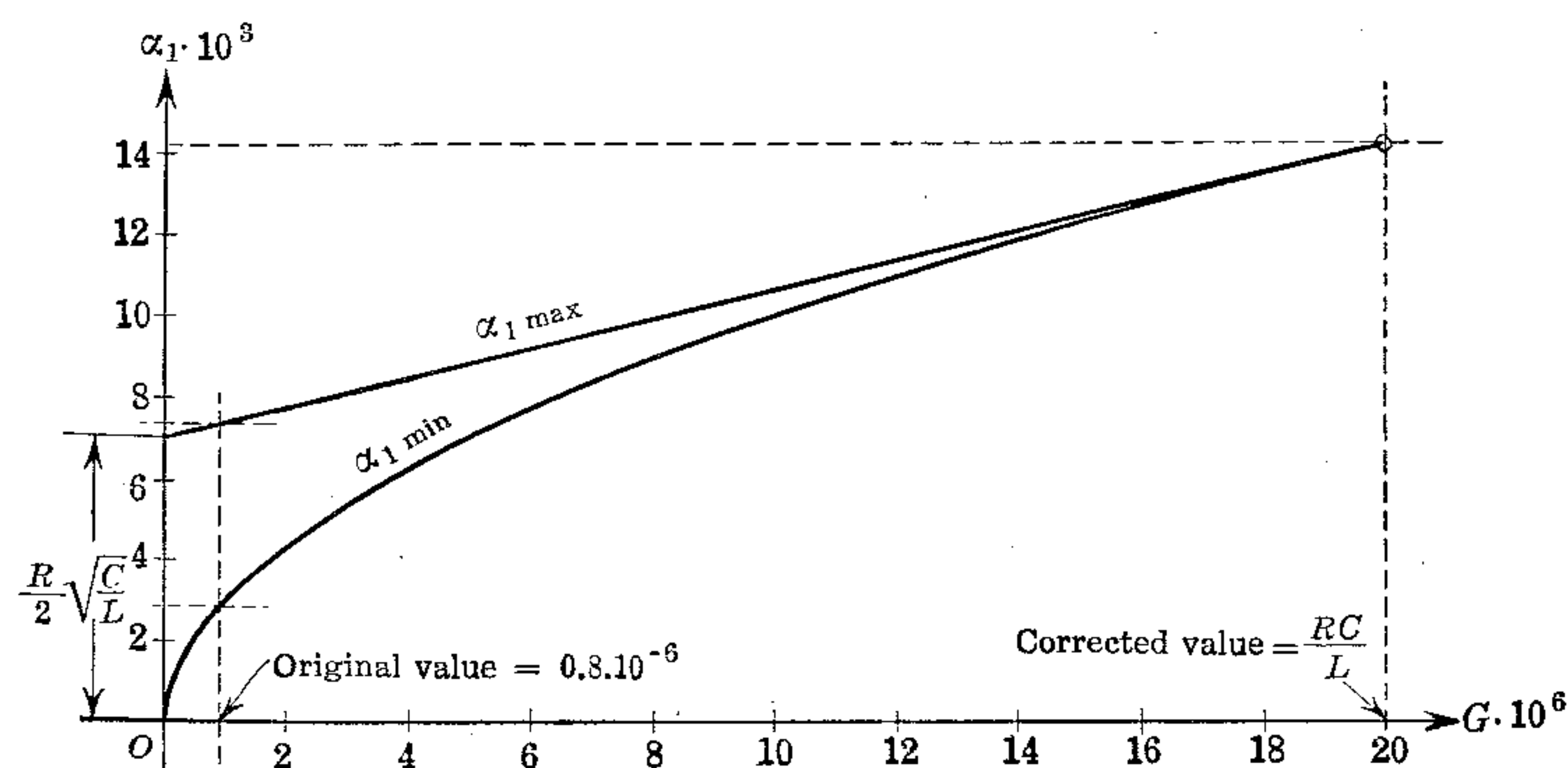


FIG. 18.—Dependence of the maximum and minimum values of attenuation upon the conductance parameter for constant values of R , L , and C .

approach the distortionless condition. Both the maximum and minimum values of the attenuation function are very materially increased by this procedure. The following table summarizes the situation with regard to the comparative results obtained from parameter readjustments.

ATHERFORCE

	Parameter Values			Attenuation Values x 10 ⁴		
	Original	30% Corrected	100% Corrected	Original	30% Corrected	100% Corrected
<i>R</i>	10	7.12	0.4	28.3-73.5	23.9-53.2	5.656
<i>L</i>	0.004	0.0328	0.1	"	28.3-32.8	28.3
<i>G</i>	0.8 x 10 ⁻⁶	6.56 x 10 ⁻⁶	20 x 10 ⁻⁶	"	81.0-93.9	141.4

5. Tolerances and frequency limits. We have just seen that a considerable improvement with regard to amplitude and phase distortion may be secured with only a partial readjustment in the line parameters. In practice it is usually possible to tolerate a certain amount of distortion depending upon the quality of service required from the facility in question. In such a case, the maximum tolerance will be specified, and the question arises as to the extent to which the parameter values must be readjusted in order to meet this specification. A glance at the functions $f_1(x)$ and $f_2(x)$ of Fig. 13 shows that the frequency range for which the facility is to be used is also essential in deciding this question. Thus, for the same maximum tolerance, a range lying above a frequency corresponding to, say, $x = 4$, will evidently require less readjustment than one of the same width lying below this value. When both the frequency range and the maximum tolerance are specified, the practical problem admits of a unique solution.

In general the process of finding the solution is not so simple as it might be on account of the fact that it involves a cut-and-try procedure. It is simplified, however, by the fact that *the average variation in the attenuation function always equals the average variation (from its ideal of the phase function.* This interesting point may be seen from the following. Let us denote the limits of the essential frequency range by x_1 and x_2 . Then the average variation of the function $f_1(x)$ within this range is

$$\frac{1}{2}\{f_1(x_2) - f_1(x_1)\},$$

while its average value throughout the same range is

$$\frac{1}{2}\{f_1(x_2) + f_1(x_1)\},$$

so that the maximum percentage variation becomes

$$\frac{f_1(x_2) - f_1(x_1)}{f_1(x_2) + f_1(x_1)} \cdot 100\%. \quad (16)$$

Similarly the percentage variation of $f_2(x)$ within this same range is

$$\frac{f_2(x_2) - f_2(x_1)}{f_2(x_2) + f_2(x_1)} \cdot 100\%. \quad (165)$$

If we substitute the relation (154) into (165), we find that it reduces to (164) except for a reversal of sign. Thus the percentage amplitude and phase distortion defined on this basis are always the same in magnitude throughout any given frequency range!

When a maximum value for either of these variations is specified together with the essential frequency range, the following procedure may be adopted for determining the corresponding set of line parameters which will meet these specifications. Let us define the magnitude of the prescribed decimal variation in either $f_1(x)$ or $f_2(x)$ as

$$\delta_\alpha = \frac{f_1(x_2) - f_1(x_1)}{f_1(x_2) + f_1(x_1)}. \quad (164a)$$

From this we may easily get

$$\frac{1 - \delta_\alpha}{1 + \delta_\alpha} = \frac{f_1(x_1)}{f_1(x_2)}. \quad (166)$$

The relation between the variable x and the angular frequency is given by the last equation (148), namely

$$x = \frac{\omega}{\sqrt{ab}}. \quad (148a)$$

The limits of the prescribed frequency range may be denoted by ω_1 and ω_2 with the restriction that $\omega_1 \leq \omega_2$. In order to determine the corresponding values of x , the product ab must be known. Since this is one of the factors which we are trying to determine, we shall have to assume various values for it. This is more convenient if we let

$$\sqrt{ab} = k\sqrt{\omega_1\omega_2}, \quad (167)$$

and then assign values to the parameter k . The square root on the right-hand side of (167), incidentally, is the geometric mean frequency of the band. It is usually denoted as

$$\omega_0 = \sqrt{\omega_1\omega_2}. \quad (168)$$

Now we can write

$$x_1 = \frac{\omega_1}{\sqrt{ab}} = \frac{1}{k}\sqrt{\frac{\omega_1}{\omega_2}}, \quad (169)$$

and

$$x_2 = \frac{\omega_2}{\sqrt{ab}} = \frac{1}{k}\sqrt{\frac{\omega_2}{\omega_1}}. \quad (170)$$

For the sake of abbreviation let

$$s = \sqrt{\frac{\omega_1}{\omega_2}} \quad (171)$$

Then

$$x_1 = \frac{s}{k}; \quad x_2 = \frac{1}{ks} \quad (172)$$

Substituting these into (166) we have

$$\frac{1 - \delta_\alpha}{1 + \delta_\alpha} = \frac{f_1\left(\frac{s}{k}\right)}{f_1\left(\frac{1}{ks}\right)} \quad (166a)$$

If we apply (152) and (154) to the right-hand side of this equation, we find

$$\frac{f_1\left(\frac{s}{k}\right)}{f_1\left(\frac{1}{ks}\right)} = \frac{f_1(ks)}{f_1\left(\frac{k}{s}\right)} \quad (173)$$

from which we see that if a certain value of k satisfies (166a) then the reciprocal of this value will also satisfy it!

Since the form of the function $f_1(x)$ depends only upon the parameter m , as may best be seen from (150a), we may proceed to choose arbitrary values for k and graphically determine the corresponding m values which satisfy (166a). By the first relation (149), m determines the ratio of a to b , namely

$$\sqrt{\frac{a}{b}} = m \pm \sqrt{m^2 - 1}, \quad (174)$$

where the minus sign gives the reciprocal value. This together with (167) then determines a and b values which satisfy the prescribed conditions regarding tolerance and frequency limits.

Note that since k and its reciprocal both satisfy the same m and hence lead to the same a -to- b ratio, it is only necessary to choose values either larger or smaller than unity. Furthermore, one choice of k and the correspondingly determined m will give us four pairs of a and b values which are the result of either of

$$\left. \begin{aligned} \sqrt{\frac{a}{b}} &= m_i + \sqrt{m_i^2 - 1} \\ \sqrt{\frac{a}{b}} &= m_i - \sqrt{m_i^2 - 1} \end{aligned} \right\} \quad (174)$$

combined with either of

$$\left. \begin{aligned} \sqrt{ab} &= k_i \omega_0 \\ \sqrt{ab} &= \frac{\omega_0}{k_i} \end{aligned} \right\} \quad (167a)$$

where k_i and m_i denote corresponding values.

The determination of m for a choice of k by means of (166a) is more conveniently accomplished if we make use of the function

$$F_1(x) = \frac{f_1(x)}{\sqrt{m}} \quad (175)$$

Referring to Fig. 13 we see that $F_1(x)$ has the property that

$$F_1(1) = 1, \quad (176)$$

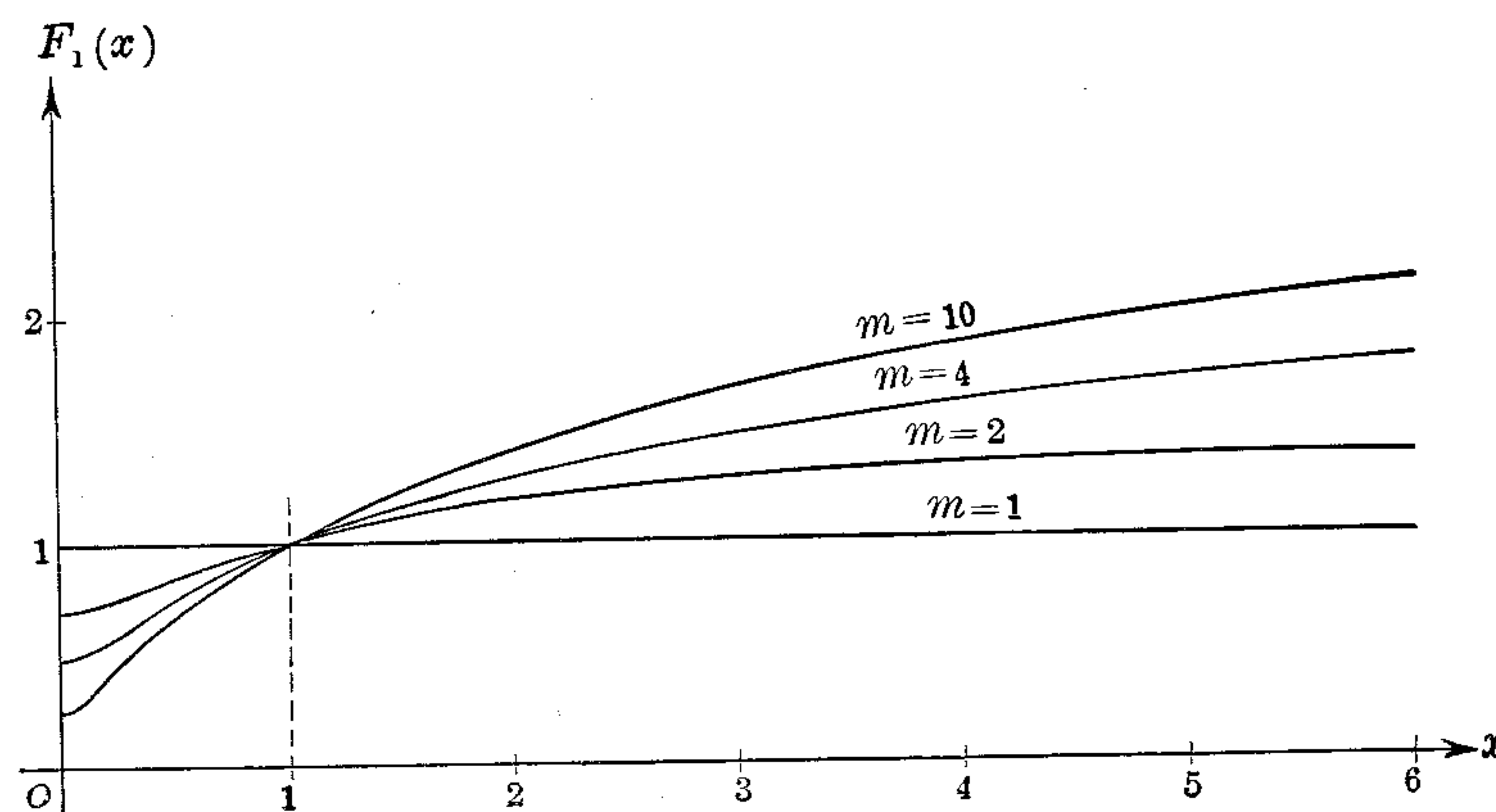


FIG. 19.—Normalized attenuation departure function, eq. (175), plotted against the frequency variable x , eq. (148), for various values of the parameter m , eq. (149).

i.e., a family of such curves for various m -values all pass through the point (1, 1). This makes a more easily usable chart. Fig. 19 shows such a family of curves.

Instead of (166a) we may, therefore, use

$$\frac{1 - \delta_\alpha}{1 + \delta_\alpha} = \frac{F_1(ks)}{F_1\left(\frac{k}{s}\right)} \quad (166b)$$

For a given s , as determined by (171), the procedure is to choose successive k values, as for example

$$k_i = 1, 2, 3, \dots$$

With each of these the chart of Fig. 19 may be entered and correspond-

ATHERFORCE

ing values of m_i found by trial which satisfy (166b) for the prescribed tolerance δ_α . The corresponding values of a and b are then found from (174a) and (167a).

With a sufficient number of these we may draw a set of curves of a versus b . Fig. 20 shows several such sets for various values of δ_α . We see that two curves, symmetrical about the 45° line, are obtained for each $\delta_\alpha \neq 0$, while the 45° line itself represents the condition for $\delta_\alpha = 0$. This might have been expected since the condition $a = b$ represents the distortionless line. The four pairs of values for a and b obtained from (174a) and (167a) for a single set of k_i and m_i give rise to four symmetrical points such as those marked 1, 2, 3, 4 in Fig. 20, two lying on one branch and two on the other.

The significance of the two branches for a given δ_α is that for each value of b there are *two* values of a which satisfy the prescribed data,

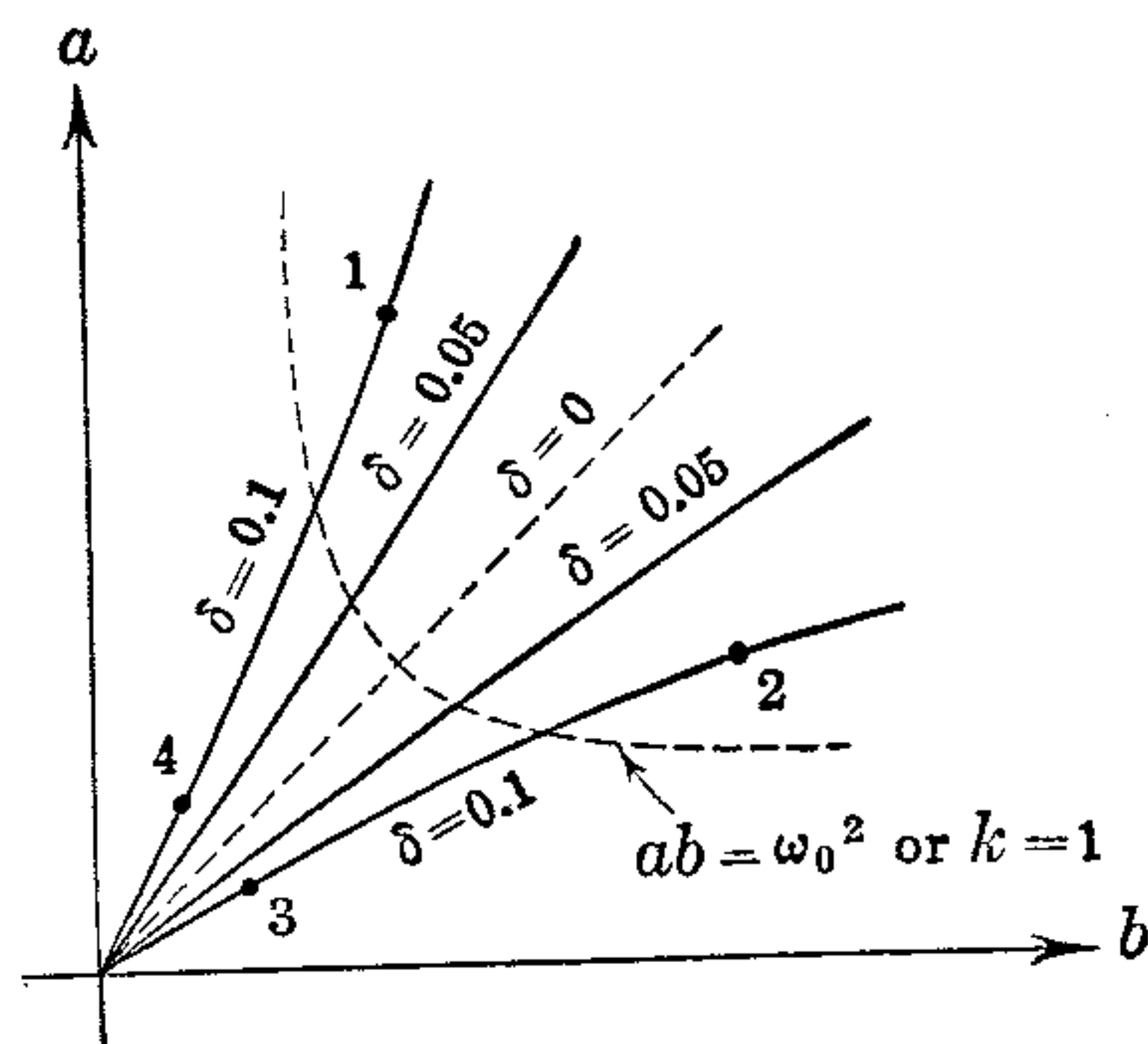


FIG. 20.—Ratio $a=R/L$ plotted against the ratio $b=G/C$ for various values of tolerance δ_α over a given frequency band.

one being larger and the other smaller than b . In any practical case we saw that for the original line or cable $a > b$, and the corrective procedure consists in decreasing a by either decreasing the resistance or increasing the inductance. This would, of course, be carried only so far as necessary. To carry the corrective procedure beyond the condition $a = b$ would be nonsense. Hence in practice only the curves of Fig. 20 lying *above* the 45° line are significant. Since b is usually fixed by the leakage and capacitance of the line, the

solution to our corrective procedure becomes unique, although theoretically a *double infinity* of solutions present themselves.

For the sake of theoretical interest it might be added that the curves of Fig. 20 will in general be grouped more closely about the 45° line for wider frequency limits and for smaller values of the geometric mean frequency ω_0 . For solutions corresponding to points below the hyperbola $k = 1$, the frequency band is located so that ω_0 lies *above* the point $x = 1$, while a point above this hyperbola places ω_0 *below* $x = 1$. Mathematically expressed this is the same as saying that for points below the hyperbola $k = 1$

$$\frac{\omega_0}{\sqrt{ab}} > 1,$$

while for those above this curve

$$\frac{\omega_0}{\sqrt{ab}} < 1.$$

For solutions lying on this hyperbola the frequency band is geometrically located about the frequency corresponding to $x = 1$.

In most practical cases conditions are such that approximations can be made which materially simplify the above general procedure. Usually the upper frequency limit ω_2 is high enough so that the function $f_1(x)$ may be assumed equal to its asymptotic value at this limit with sufficient accuracy. This means that in (166) we may let

$$f_1(x_2) = m. \quad (177)$$

For the value of the function at the lower limit it is sufficient to use the first two terms of the expansion (159). For values of δ_α less than 0.1, the error introduced by this procedure is not very great as may easily be verified. Consistent with these approximations for the right-hand side of (166), the left-hand side may likewise be put into an approximate form which is justified for values of δ_α less than 0.1.

Thus

$$\frac{1 - \delta_\alpha}{1 + \delta_\alpha} \cong (1 - \delta_\alpha)^2 \cong 1 - 2\delta_\alpha. \quad (178)$$

With these simplifications (166) becomes

$$1 - 2\delta_\alpha = 1 - \frac{n^2}{2x_1^2}. \quad (179)$$

Expressing this in terms of frequency by means of (148a) we have

$$\delta_\alpha = \frac{n^2 ab}{4\omega_1^2},$$

or by using the second relation (149) this is

$$\delta_\alpha = \frac{(a - b)^2}{16\omega_1^2}. \quad (180)$$

When the tolerance and the lower frequency limit are specified, the condition on the line parameters is expressed by

$$(a - b) = 4\omega_1 \sqrt{\delta_\alpha}. \quad (180a)$$

Substituting from (148) we have

$$\frac{R}{L} - \frac{G}{C} = 4\omega_1 \sqrt{\delta_\alpha}. \quad (180b)$$

ETHERFORCE

If the ratio G/C is given, the necessary R/L ratio for which the variation δ_α will not be exceeded by either $f_1(x)$ or $f_2(x)$ for all frequencies above ω_1 may easily be found. As an illustration suppose we consider the open line represented by the parameter values on page 90 and stipulate that the inductance shall be increased until the maximum average variation δ_α does not exceed 0.05 for all frequencies in the voice range. Assuming that the lowest essential voice frequency is 100 cycles per second, we have

$$\omega_1 = 2\pi \cdot 100.$$

The upper frequency limit need not be considered since we have assumed it to be sufficiently high so that $f_1(x)$ very nearly equals the value m at this limit. The accuracy of this assumption may later be checked. We thus have

$$\left(\frac{R}{L}\right)_{\text{new}} - \frac{G}{C} = 8\pi \cdot 100 \cdot \sqrt{0.05} = 562,$$

and since for this line

$$\frac{G}{C} = 100,$$

we get

$$\left(\frac{R}{L}\right)_{\text{new}} = 662$$

for the required ratio of resistance to inductance. Since the given ratio is

$$\left(\frac{R}{L}\right)_{\text{old}} = 2500,$$

we see that the inductance must be increased until

$$L_{\text{new}} = 3.78L_{\text{old}}.$$

In practice the introduction of additional inductance also increases the resistance. The process of taking this into account is simple and need not be taken up here. With this increase the inductance of the loaded line will be

$$L_{\text{new}} = 0.0151 \text{ henry per mile.}$$

If we assume that the upper limit of the voice frequency range is

$$\omega_2 = 2\pi \cdot 4000,$$

we find by substituting into (159) that

$$f_1(x_2) = m(1 - 0.000062)$$

which is certainly near enough to the value m . The result (180b) is, therefore, a convenient criterion. For larger values of δ_α it will indicate an R/L ratio somewhat smaller than necessary, but for most practical purposes its accuracy is sufficient. In cases of doubt, the error can, of course, be determined.

6. Phase and group velocity. In this section we wish to point out certain transmission properties of the long line which are brought out by a closer study of the phase function α_2 . In the previous chapter we saw by equation (80a) p. 50 that the velocity of phase propagation, which is the velocity with which the waves travel in the harmonic steady state, is given by

$$v = \frac{\omega}{\alpha_2}. \quad (80a)$$

If we use (147b) this is

$$v = \frac{1}{\sqrt{LC} \cdot f_2(x)}. \quad (181)$$

In the foregoing we have seen that when the line is dissipationless, or satisfies the condition (160) which makes it distortionless, the function $f_2(x)$ becomes unity at all frequencies. Since in general $f_2(x)$ is greater than unity, we see that in the dissipationless or distortionless cases the phase velocity is constant and has its maximum value, which is

$$v_0 = \frac{1}{\sqrt{LC}}. \quad (182)$$

In terms of this value, and using the relation (154), we may write instead of (181)

$$v = v_0 \cdot \frac{f_1(x)}{m}. \quad (181a)$$

This clearly shows the variation of phase velocity with frequency in the general case. In fact the curve for $f_1(x)$ of Fig. 13 may be used to represent the velocity v by making a suitable change in the vertical scale in agreement with (181a). We thus see that the lower frequencies travel more slowly and that the value v_0 (which we recall is nearly equal to the velocity of light for open lines) is approached asymptotically as the frequency tends toward infinity. The ratio of this highest velocity v_0 to the lowest is, in the general case, equal to m .

This is the situation in the steady state when a single harmonic voltage is impressed upon the line. In communication work, however, we rarely find a situation as simple as this. In fact it may be questionable whether the average condition in practice may even be considered as

ATHERFORCE

a steady state. However, we shall leave this question out of the argument until a later chapter. Meanwhile it is quite instructive as well as interesting to consider what happens when a steady state is composed of a linear superposition of a number of harmonic functions of different frequencies. Such an array of functions we shall call a **frequency group**. Furthermore, we shall confine our attention to a particular kind of frequency group, namely one for which all the amplitudes are equal and all the frequency differences equal.

Such a frequency group is illustrated in Fig. 21, which, incidentally, is called an **amplitude spectrum** representation for the group. The abscissae for this figure indicate the angular frequencies of the individual components; the ordinates indicate their respective amplitudes. This group is composed of one component whose frequency is the arithmetic mean, and n adjacent pairs equally spaced. The frequency interval is denoted as $\delta\omega$, and the "width" of the group is designated as

$$\Delta\omega = (2n + 1)\delta\omega. \quad (183)$$

It may seem to the reader that this group is rather special and, therefore, would hardly have any practical significance. However, an arbitrary distribution of amplitude versus frequency can always be approximately thought of as composed of a superposition of such groups as these. Such a representation for an arbitrary amplitude versus frequency distribution will become more exact the smaller the width $\Delta\omega$ of the individual groups is chosen. We will see later that by letting the group width approach zero, any time function may be accurately represented as a linear superposition of an infinite number of such groups. Thus the consideration of a group as is illustrated by Fig. 21 forms the basis for the exact analysis of the most general case.

To continue with the consideration of this frequency group, our first object is to show what the resulting wave picture for such an array of frequencies looks like. In order to do this we must write down the analytic expression for a group of waves whose frequencies and amplitudes correspond to the above distribution. If we assume for simplicity that the amplitudes are all unity, then the analytic expression for a component wave of angular frequency ω is

$$\cos(\alpha_2 x - \omega t).^1$$

¹ The variable x here represents distance along the line as in Chapter II.

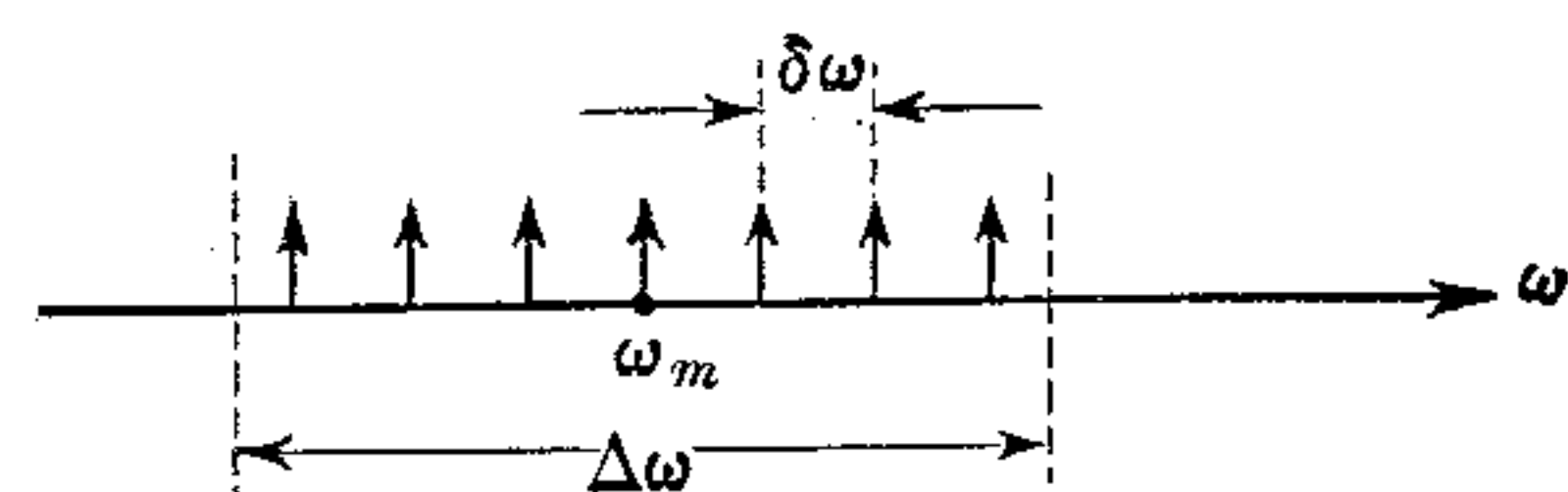


FIG. 21.—Amplitude spectrum of a uniform frequency group.

In particular for the *mean* frequency ω_m this is

$$\cos(\alpha_{2m}x - \omega_m t).$$

If we denote the corresponding change in α_2 to a frequency change $\delta\omega$ by $\delta\alpha_2$, then the next component wave in the group is given by

$$\cos[(\alpha_{2m} + \delta\alpha_2)x - (\omega_m + \delta\omega)t].$$

In order to simplify the writing of the entire group, let us introduce the notation

$$\left. \begin{aligned} \Omega &= \alpha_2 x - \omega t \\ \delta\Omega &= \delta\alpha_2 x - \delta\omega t \\ \Omega_m &= \alpha_{2m} x - \omega_m t \end{aligned} \right\}. \quad (184)$$

Then the wave group corresponding to the frequency group defined above becomes

$$f(x - vt) = \cos \Omega_m t + \left\{ \begin{aligned} &\cos(\Omega_m + \delta\Omega) + \cos(\Omega_m + 2\delta\Omega) + \cdots + \cos(\Omega_m + n\delta\Omega) \\ &\cos(\Omega_m - \delta\Omega) + \cos(\Omega_m - 2\delta\Omega) + \cdots + \cos(\Omega_m - n\delta\Omega) \end{aligned} \right\}. \quad (185)$$

This finite sum of harmonic components may be evaluated by applying the trigonometric formulae for sums and differences of angles.¹ Thus (185) may be written

$$f(x - vt) = (2n + 1) \cdot F(\delta\Omega) \cdot \cos \Omega_m, \quad (185a)$$

where

$$F(\delta\Omega) = \frac{\sin \left[(2n + 1) \frac{\delta\Omega}{2} \right]}{(2n + 1) \cdot \sin \frac{\delta\Omega}{2}}. \quad (185b)$$

The result (185a) shows that the wave group may be visualized as consisting of the mean-frequency wave

$$\cos \Omega_m = \cos(\alpha_{2m}x - \omega_m t)$$

enclosed by an envelope whose form is determined by

$$(2n + 1) \cdot F(\delta\Omega).$$

This envelope is itself a wave; i.e., it has a fixed form but travels along the line as time increases. In order to get a better idea of what the shape of this envelope looks like, we will plot it versus distance for the instant $t = 0$.

¹ See for example, W. E. Johnson, "A Treatise on Trigonometry," Macmillan & Co., 1889, p. 320, e.s.

Then we have

$$F(\delta\Omega) = F(\delta\alpha_2 x) = \frac{\sin \left[(2n+1) \cdot \frac{\delta\alpha_2}{2} \cdot x \right]}{(2n+1) \cdot \sin \frac{\delta\alpha_2}{2} x} \quad (186)$$

If the number of components in the group is large, i.e., if n is large, then for small values of x this is approximately given by

$$F(\delta\alpha_2 x) \cong \frac{\sin \left[(2n+1) \cdot \frac{\delta\alpha_2}{2} \cdot x \right]}{(2n+1) \cdot \frac{\delta\alpha_2}{2} \cdot x} \quad (186a)$$

which is of the form

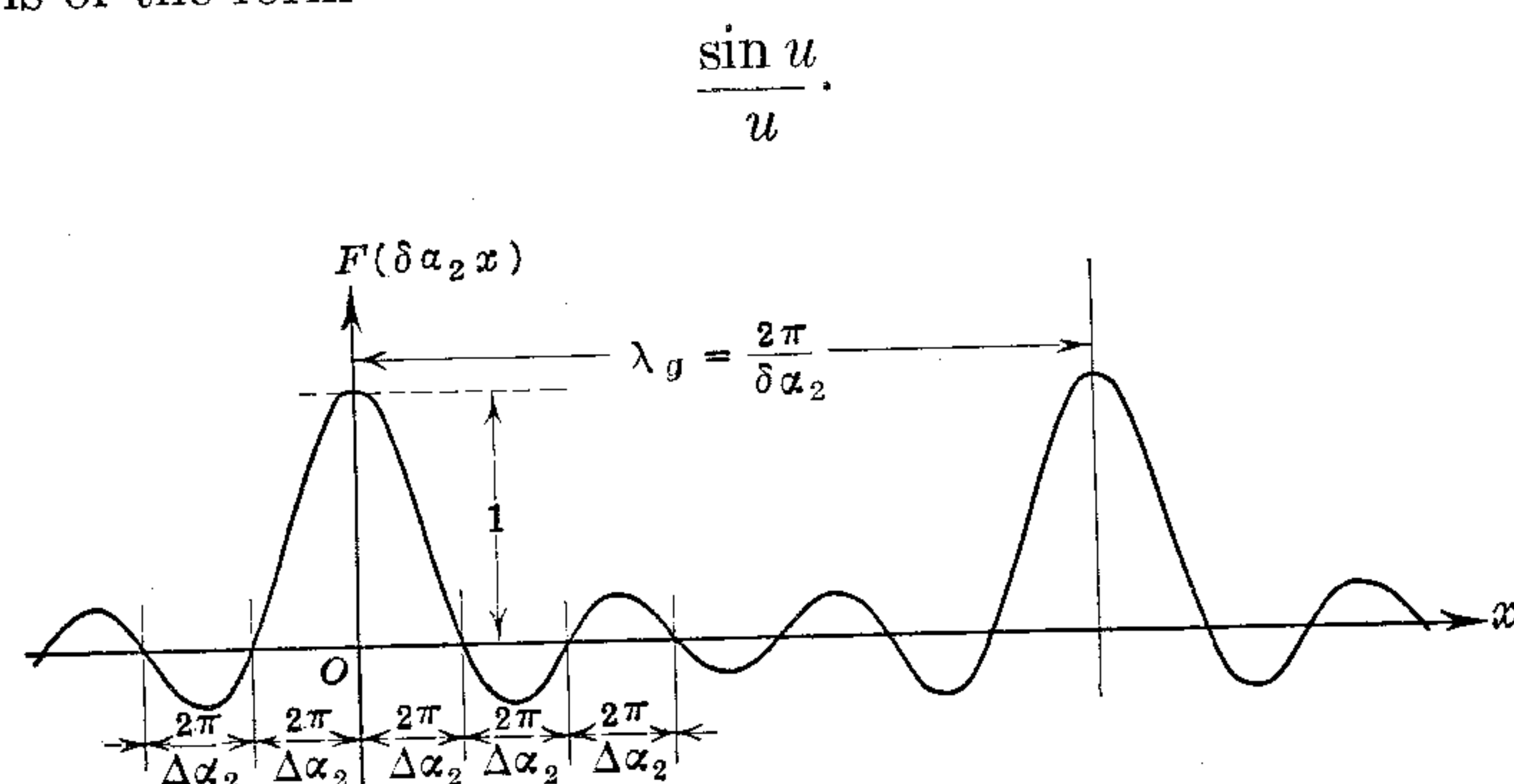


FIG. 22.—Wave envelope corresponding to the frequency group of Fig. 21.

This function crops up quite frequently in connection with network problems, and the student may as well familiarize himself somewhat with its general characteristics. The form (186a), of course, may be used to plot the function only in the vicinity of $x = 0$. The actual function (186) is obviously periodic in x ; i.e., it repeats after a distance

$$\lambda_g = \frac{2\pi}{\delta\alpha_2} \quad (187)$$

In Fig. 22 is shown a plot of (186) for $n = 3$. In this figure the quantity

$$\Delta\alpha_2 = (2n+1)\delta\alpha_2 \quad (188)$$

is used to denote the *width* of the group with respect to α_2 just as $\Delta\omega$ in (183) denotes the *width* with respect to ω .

Thus we see that the envelope which encloses the mean-frequency wave has the effect of grouping or “bunching” this wave into humps which come at intervals of λ_g miles. This distance, which is inversely

proportional to the increment $\delta\alpha_2$ and hence inversely proportional to the corresponding frequency increment $\delta\omega$, is called the **group wave length**. The length of any “ripple” in the envelope, which is half the length of the main hump, is

$$\frac{2\pi}{\Delta\alpha_2} = \frac{2\pi}{(2n+1) \cdot \delta\alpha_2}$$

and hence depends upon the “width” of the frequency group.

The complete picture of the wave group as it travels along the line is, therefore, as illustrated by Fig. 23, which is again drawn for $n = 3$. A further point of interest in connection with this wave group is now revealed by the fact that the velocity with which the envelope travels

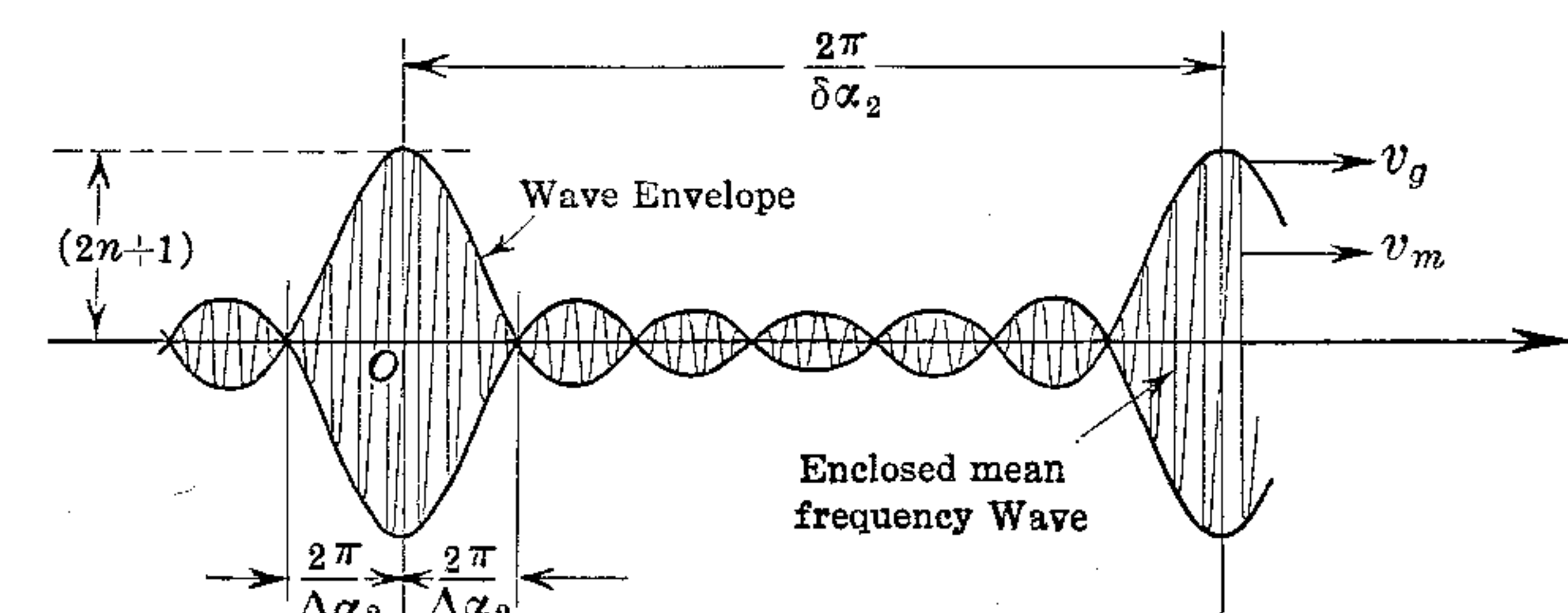


FIG. 23.—Resulting wave group corresponding to the frequency group of Fig. 21.

is in general not equal to that with which the enclosed mean-frequency wave travels. The latter is evidently

$$(v_{ph})_m = \frac{\omega_m}{\alpha_{2m}} \quad (189)$$

Since the arguments of the sine-functions in (185b) involve the quantity

$$\delta\Omega = \delta\alpha_2 x - \delta\omega t,$$

a point of constant phase for the envelope will be given by

$$d(\delta\alpha_2 x - \delta\omega t) = \delta\alpha_2 \cdot dx - \delta\omega \cdot dt = 0,$$

from which

$$\frac{dx}{dt} = v_g = \frac{\delta\omega}{\delta\alpha_2} \quad (190)$$

This is the velocity with which the *envelope* travels. If the width $\Delta\omega$ of the frequency group is appreciable, the value

$$\frac{\delta\omega}{\delta\alpha_2}$$

may vary somewhat throughout the group. In general, the group

ATHERFORCE

width $\Delta\omega$ will be taken small enough so that this variation may be neglected, and we can write

$$v_g = \left(\frac{d\omega}{d\alpha_2} \right)_{\omega = \omega_m} \quad (190a)$$

That is, the velocity of the envelope, which is also called the **group velocity**, is equal to the inverse slope of the α_2 -versus- ω curve at the point $\omega = \omega_m$.

In the dissipationless or the distortionless cases for which

$$\alpha_2 = \omega\sqrt{LC},$$

$$v_{ph} = v_g = \frac{1}{\sqrt{LC}},$$

we have

i.e., both the *phase* and *group velocities* are constant and equal to each other. In general, however, the group and phase velocities will vary with frequency and will not equal each other.

In order to study the variation of the group velocity with frequency for the long line, we must first determine the variation of the derivative of α_2 with respect to ω . The expression for this derivative becomes quite involved if written entirely in terms of ω . However, if by the use of (148) and (147b) we first write

$$\alpha_2 = \sqrt{RG} \cdot x \cdot f_2(x),^1$$

and then note that

$$\frac{d\alpha_2}{d\omega} = \frac{d\alpha_2}{dx} \cdot \sqrt{\frac{LC}{RG}},$$

so that

$$\frac{d\alpha_2}{d\omega} = \sqrt{LC} \cdot \frac{d}{dx}(x \cdot f_2(x)),$$

we get by using (151a) and recalling the substitution (153)

$$\frac{d\alpha_2}{d\omega} = \sqrt{LC} \left[\frac{(\sqrt{m^2 + y^2} + y)^{1/2} \cdot (\sqrt{m^2 + y^2} + y + \frac{1}{x})}{2\sqrt{x}\sqrt{m^2 + y^2}} \right]. \quad (191)$$

This form is quite convenient for plotting. Fig. 24 shows this derivative as a function of x for $m = 4$. The units for the vertical scale are equal to $\sqrt{LC}$, which is the dissipationless or distortionless value. We see that the function has a minimum value less than this, which is quite an interesting point. In order to locate the frequency at which this minimum occurs, we must equate the second derivative of α_2 to zero and solve for the corresponding x . This process becomes exceedingly

¹ Note that here x is the variable as defined by (148) and not distance.

involved, as may easily be appreciated. The simplest way of accomplishing it is not to attempt differentiating (191) again, but to go back to the original expression (142) for α . Since the derivative of the imaginary part equals the imaginary part of the derivative, we may accomplish the desired end by differentiating (142) twice with respect to ω ,

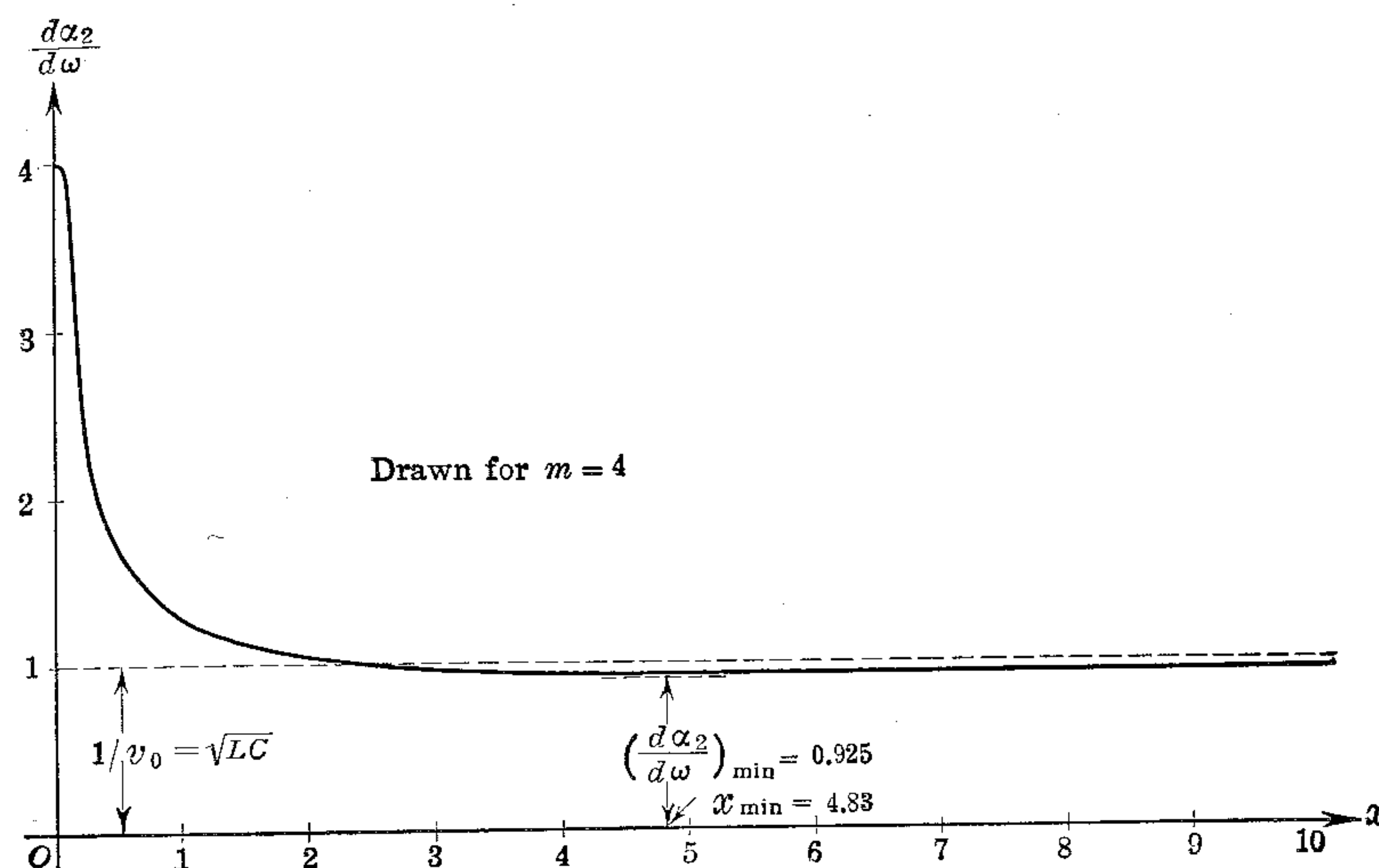


FIG. 24.—Delay time (units equal to $\sqrt{LC}$ seconds) of the uniform transmission line as a function of the frequency variable x , eq. 148, for a constant value of the parameter m , eq. 149, assuming constant line parameters.

and then equate the imaginary part of the result to zero. We shall omit the actual work here, and merely state the result, namely that the minimum of (191) occurs at

$$x_{\min} = \frac{\sqrt{m^2 + 3} + m}{\sqrt{3}}, \quad (192)$$

and that the minimum value found by substituting this back into (191) becomes

$$\left(\frac{d\alpha_2}{d\omega} \right)_{\min} = \sqrt{LC} \cdot \frac{\sqrt{3}}{4} \left\{ \sqrt{\frac{m}{\sqrt{m^2 + 3} + m}} + \sqrt{\frac{\sqrt{m^2 + 3} + m}{m}} \right\}. \quad (193)$$

It is now interesting to compare the phase and group velocities as functions of frequency. They are respectively

$$\left. \begin{aligned} v_{ph} &= \frac{\omega}{\alpha_2} \\ v_g &= \frac{d\omega}{d\alpha_2} \end{aligned} \right\}. \quad (194)$$

Using (181a) for the former and inverting Fig. 24 for the latter, we find the curves of Fig. 25 for these two quantities, again drawn for $m = 4$. We see from this figure that in general the group velocity is *always* larger than the phase velocity! This means that when a wave group travels along a line, the velocity of the envelope, i.e., of the groups or humps, will in general be greater than that of the enclosed wave, and that these velocities merge toward the distortionless value v_0 as the mean frequency of the group tends toward infinity.

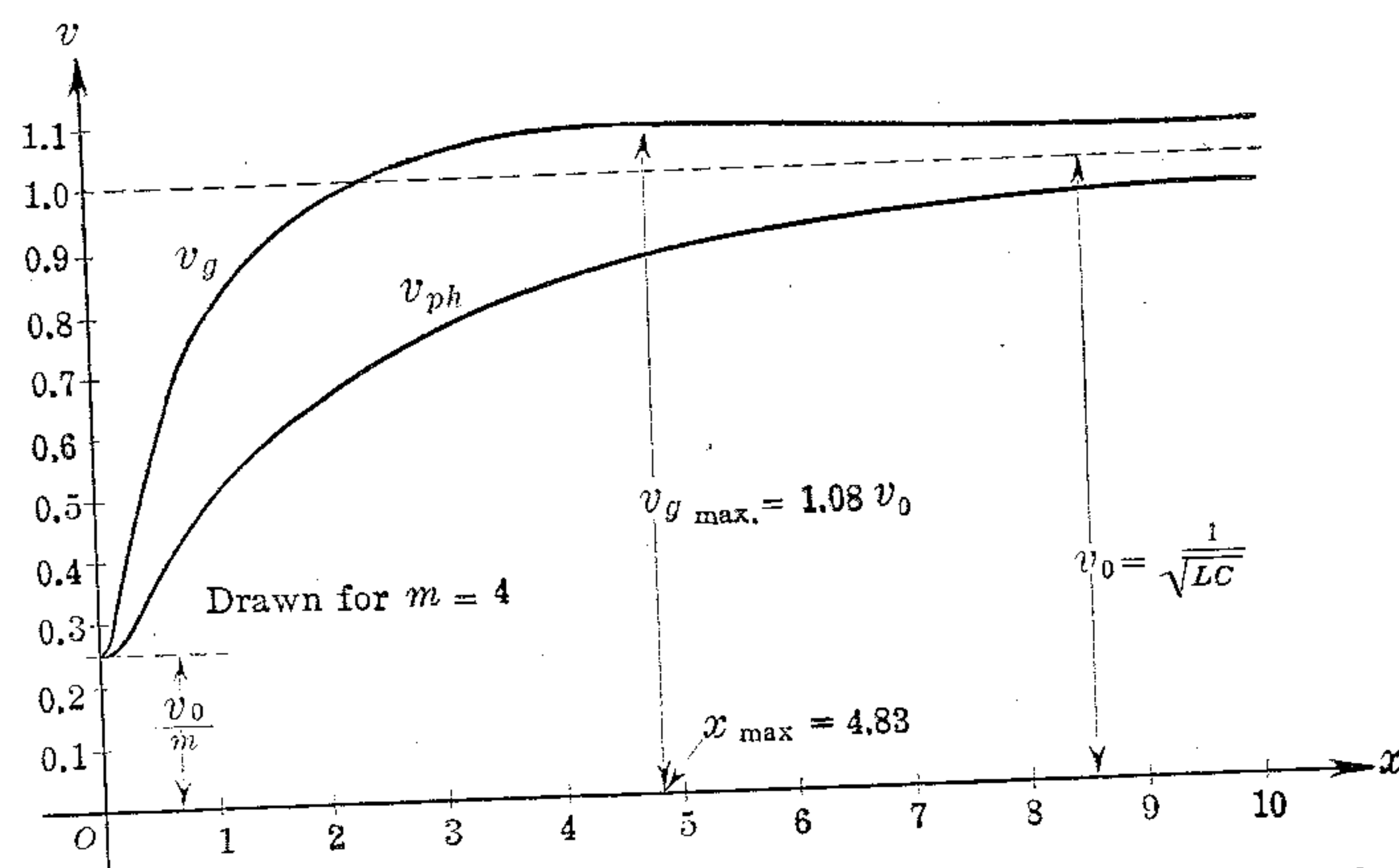


FIG. 25.—Phase and group velocities of the uniform transmission line as functions of the frequency variable x . These curves correspond to the conditions underlying Fig. 24.

From (193) we note that this expression has its smallest value for $m \rightarrow \infty$, and that this value is

$$\left[\left(\frac{d\alpha_2}{d\omega} \right)_{\min} \right]_{m \rightarrow \infty} = \sqrt{LC} \cdot \frac{3\sqrt{3}}{4\sqrt{2}}. \quad (195)$$

This means that the largest maximum value of v_g will occur for lines for which either

$$\frac{R}{L} \gg \frac{G}{C}$$

or

$$\frac{G}{C} \gg \frac{R}{L},$$

and that this largest value is

$$v_{g \max \max} = \frac{4\sqrt{2}}{3\sqrt{3}} v_0 = 1.088 v_0. \quad (196)$$

7. Discussion of the characteristic impedance function. This function, like the propagation function, is also in general complex, and hence may be split into real and imaginary components. However, it is usually more convenient to study the characteristic impedance function in its polar form, i.e., in the form of magnitude and angle. Thus we write

$$Z_0 = \sqrt{\frac{R + jL\omega}{G + jC\omega}} = |Z_0| \angle \zeta \quad (197)$$

and proceed to determine the functions $|Z_0|$ and ζ . For the magnitude we evidently have

$$|Z_0| = \left(\frac{R^2 + L^2\omega^2}{G^2 + C^2\omega^2} \right)^{1/4} = \sqrt{\frac{L}{C}} \cdot \left(\frac{a^2 + \omega^2}{b^2 + \omega^2} \right)^{1/4}, \quad (198)$$

where the relations (148) have been used. The quantity $\sqrt{L/C}$ is obviously the dissipationless value. The following parenthesis factor in (198), therefore, shows the effect of dissipation and hence contains all that is of interest in the study of $|Z_0|$. Let us write

$$|Z_0| = \sqrt{\frac{L}{C}} \cdot z_0, \quad (198a)$$

and then with the last relation (148) we have

$$z_0(x) = \left(\frac{\frac{a}{b} + x^2}{\frac{b}{a} + x^2} \right)^{1/4}. \quad (199)$$

From this result we can obtain the interesting relation that

$$z_0(x) \cdot z_0\left(\frac{1}{x}\right) = \sqrt{\frac{a}{b}} \quad (200)$$

from which we note that when $z_0(x)$ has been plotted for the interval $0 < x < 1$, it may be extended into the interval $1 < x < \infty$ by geometrical construction according to (200). From this relation it also follows that

$$z_0(1) = \sqrt[4]{\frac{a}{b}}, \quad (201)$$

as is also evident from (199).

In the vicinity of $x = 0$ the relation (199) may be expanded to obtain

$$\begin{aligned} z_0(x) &= \left(\frac{a}{b} + x^2\right)^{1/4} \left(\frac{b}{a} + x^2\right)^{-1/4} = \sqrt{\frac{a}{b}} \left(1 + \frac{bx^2}{a}\right)^{1/4} \left(1 + \frac{ax^2}{b}\right)^{-1/4} \\ &= \sqrt{\frac{a}{b}} \left(1 + \frac{bx^2}{4a} + \dots\right) \left(1 - \frac{ax^2}{4b} + \dots\right) \\ &= \sqrt{\frac{a}{b}} \left\{1 - \frac{x^2}{4} \left(\frac{a}{b} - \frac{b}{a}\right) + \dots\right\}, \end{aligned} \quad (202)$$

while for large values of x we may write

$$\begin{aligned} z_0(x) &= \left(1 + \frac{a}{bx^2}\right)^{1/4} \left(1 + \frac{b}{ax^2}\right)^{-1/4} \\ &= \left(1 + \frac{a}{4bx^2} + \dots\right) \left(1 - \frac{b}{4ax^2} + \dots\right) \\ &= 1 + \frac{1}{4x^2} \left(\frac{a}{b} - \frac{b}{a}\right) + \dots \end{aligned} \quad (203)$$

Thus we see that $z_0(x)$ varies between the value $\sqrt{a/b}$ at zero to the asymptotic value unity at infinity. The value (201) at $x = 1$ is thus the geometric mean. The slope at this point is found from differentiating (199) and is

$$\left(\frac{dz_0}{dx}\right)_{x=1} = -\frac{1}{2} \sqrt{\frac{a}{b}} \cdot \left(\frac{a-b}{a+b}\right). \quad (204)$$

With this information it is a simple matter to plot $z_0(x)$ for any given case. This plot will depend only upon the ratio (a/b) , and hence one such plot may cover a number of cases.

So far we have said nothing about the angle $\zeta(\omega)$ appearing in (197). This is evidently expressible as

$$\zeta(\omega) = \frac{1}{2} \{ \angle(R + jL\omega) + \angle(G - jC\omega) \} \quad (205)$$

where the sign $\angle$ is used to indicate the words "angle of." This expression may be put into the form

$$\begin{aligned} \zeta(\omega) &= \frac{1}{2} \angle(R + jL\omega)(G - jC\omega) \\ &= \frac{1}{2} \angle\{RG + LC\omega^2 + j\omega(LG - RC)\} \\ &= -\frac{1}{2} \tan^{-1} \left\{ \frac{(RC - LG)\omega}{RG + LC\omega^2} \right\}, \end{aligned}$$

which, with the help of (148) and (149), may be written

$$\zeta(x) = -\frac{1}{2} \tan^{-1} \left(\frac{2nx}{1+x^2} \right). \quad (206)$$

The most obvious characteristic of this function is the fact that

$$\zeta\left(\frac{1}{x}\right) = \zeta(x), \quad (207)$$

which will be useful in plotting. For small values of x we have

$$\zeta(x) \cong -\frac{1}{2} \tan^{-1} 2nx \cong -nx, \quad (207a)$$

from which we see that

$$\left. \begin{aligned} \zeta(0) &= 0 \\ \left(\frac{d\zeta}{dx}\right)_{x=0} &= -n \end{aligned} \right\}. \quad (208)$$

For large values of x , on the other hand,

$$\zeta(x) \cong -\frac{1}{2} \tan^{-1} \frac{2n}{x} \cong -\frac{n}{x}, \quad (207b)$$

which shows that

$$\zeta(\infty) = 0.$$

It is thus quite evident that $\zeta(x)$ has a maximum value somewhere between zero and infinity. This may be determined by the usual process of setting the derivative equal to zero. We find

$$\frac{d\zeta}{dx} = \frac{n(x^2 - 1)}{1 + 2(m^2 + n^2)x^2 + x^4} = 0,$$

from which the maximum occurs for

$$x_{\max} = \pm 1. \quad (209)$$

Substituting this back into (206) we have

$$\zeta_{\max} = \mp \frac{1}{2} \tan^{-1} n. \quad (210)$$

These results regarding the functions $z_0(x)$ and $\zeta(x)$ are summarized in the plot of Fig. 26 which is drawn for $m = 4$ and $a > b$. Here, as with the propagation function, we see that the ideal behavior will be obtained for the condition

$$a = b; \frac{R}{L} = \frac{G}{C},$$

for which obviously

$$\left. \begin{aligned} |Z_0| &= \sqrt{\frac{L}{C}} \\ \zeta &= 0 \end{aligned} \right\}, \quad (211)$$

i.e., the characteristic impedance reduces to a real constant, equal in value to that which obtains for the non-dissipative case.

For actual lines and cables the question arises as to how the param-

ATHERFORCE

eters must be altered in order that the maximum variations in $|Z_0|$ and ζ may not exceed a prescribed value over a given frequency range. This problem may be solved in an analogous fashion to that used in solving the corresponding problem in connection with the propagation function. There we saw that for most practical purposes an approximate method would answer. Here we shall, therefore, discuss only the approximate solution.

A glance at Fig. 26 reveals the fact that with the characteristic impedance function the region of maximum distortion is in general also limited to the vicinity of the origin, just as with the attenuation and phase functions. At the higher frequencies both the magnitude and

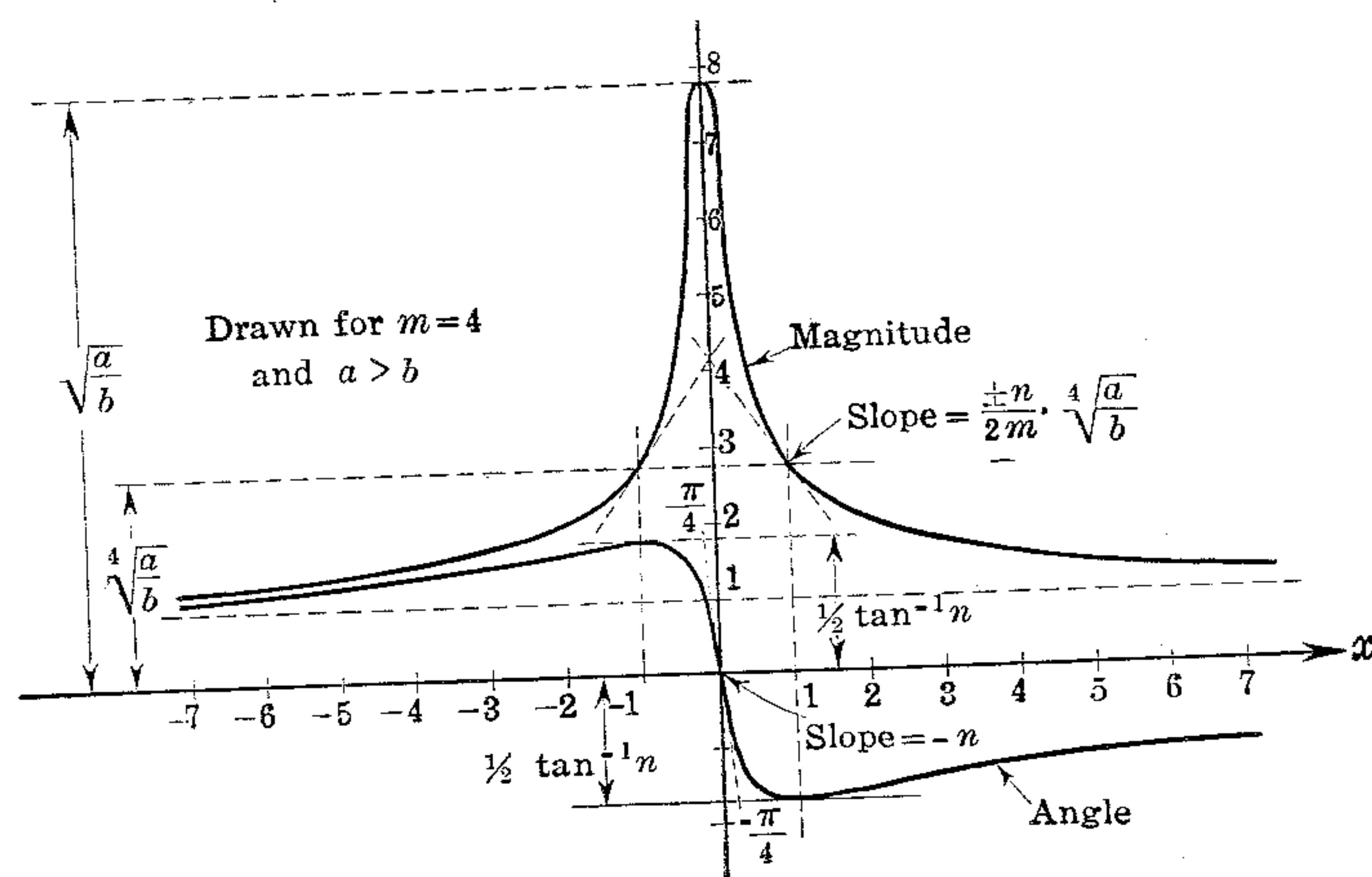


FIG. 26.—Magnitude and angle of the normalized characteristic impedance function, eqs. (197) and (198a), versus the frequency variable x .

the angle of Z_0 approach their constant asymptotic values, and hence vary but little with frequency. Thus, with the lower limit of the desired frequency range specified, the problem is to determine the relation between the line parameters for which the average variations in $z_0(x)$ and $\zeta(x)$ will stay within a prescribed tolerance.

An approximate solution to this problem is easily found if we make use of the relations (203) and (207b) for $z_0(x)$ and $\zeta(x)$, which are valid for large values of x . In the case of $z_0(x)$, the average variation over a frequency range from ω_1 to infinity is obviously half the total variation over this range. Denoting the average tolerance in $z_0(x)$ by δ_{z_0} we, therefore, have by (203)

$$\delta_{z_0} = \{z_0(\infty) - z_0(x_1)\} / \{z_0(\infty) + z_0(x_1)\} \cong -\frac{1}{8x_1^2} \left(\frac{a}{b} - \frac{b}{a} \right), \quad (212)$$

from which, with the use of (148),

$$a^2 - b^2 = 8\omega_1^2 \delta_{z_0}. \quad (212a)$$

When the lower frequency limit ω_1 and the maximum tolerance δ_{z_0} are specified, the amount by which the resistance must be reduced or the inductance increased may be found from

$$\left(\frac{R}{L} \right)_{\text{new}} = \sqrt{8\omega_1^2 \delta_{z_0} + \frac{G^2}{C^2}}, \quad (212b)$$

where it is assumed that G and C are fixed.

In the case of $\zeta(x)$ the average variation between ω_1 and ∞ equals the average value throughout this range. If we denote this average value by δ_ζ , and use (207b) we have

$$\delta_\zeta = \frac{1}{2} \zeta(x_1) = -\frac{n}{2x_1}. \quad (213)$$

Since the algebraic sign is of no consequence if δ_ζ is to indicate a magnitude only, we get by using (148) and (149)

$$a - b = 4\omega_1 \delta_\zeta. \quad (213a)$$

When ω_1 , δ_ζ , and b are given we have

$$\left(\frac{R}{L} \right)_{\text{new}} = 4\omega_1 \delta_\zeta + \frac{G}{C} \quad (213b)$$

for the determination of the required resistance to inductance ratio.

Since the average value δ_ζ will usually be small, the tangent of this angle may be replaced by the angle itself with sufficient accuracy as is also done in the relations (207a) and (207b). Hence to prescribe that δ_ζ shall be, say, 0.1 is equivalent to specifying that the average value of the imaginary part of the characteristic impedance shall not exceed 10 per cent of the magnitude of Z_0 . The quantity $\delta_\zeta \cdot 100$, therefore, indicates the average value in percentage of $|Z_0|$ which the imaginary part of Z_0 equals throughout the specified frequency range.

8. Net effect of various tolerances. Having studied the variations in the propagation and characteristic impedance functions over specified frequency ranges, the reader will probably be interested in how these variations affect the amplitude and phase of the voltage and current ratios, i.e., in the net effect of the tolerances in α and Z_0 upon the transmission of voltage and current. In this section we shall attempt to show how this situation may be studied.

In section 17 of Chapter II we derived expressions for the ratios of input to output voltage and current, as well as for the ratios of the received voltage and current without the line to the corresponding

quantities with the line inserted between the sending- and receiving-end impedances. These ratios, which are referred to as *transmission* and *insertion* ratios, respectively, we repeat here for convenience, using the same equation numbers as before

$$\frac{E_S}{E_R} = e^{\alpha l} \cdot (1 + r_R)^{-1} \cdot (1 + r_R e^{-2\alpha l}), \quad (128a)$$

$$\frac{I_S}{I_R} = e^{\alpha l} \cdot (1 - r_R)^{-1} \cdot (1 - r_R e^{-2\alpha l}), \quad (131)$$

$$\frac{E_R'}{E_R} = \frac{I_R'}{I_R} = e^{\alpha l} \cdot (1 - r_S r_R)^{-1} \cdot (1 - r_S r_R e^{-2\alpha l}). \quad (130)$$

The volt-ampere ratios are readily formed from these.

The logarithms of these ratios we shall consider as representing the *actual* or *effective* propagation functions corresponding to either the transmission or insertion bases. These we denote by the letter γ , and distinguish in the following manner:

$$\left. \begin{aligned} \gamma' &= \ln\left(\frac{E_R'}{E_R}\right) = \ln\left(\frac{I_R'}{I_R}\right) = \ln\sqrt{\frac{E_R' I_R'}{E_R I_R}}; \\ \gamma_E &= \ln\left(\frac{E_S}{E_R}\right); \gamma_I = \ln\left(\frac{I_S}{I_R}\right); \gamma_{EI} = \ln\sqrt{\frac{E_S I_S}{E_R I_R}}. \end{aligned} \right\} \quad (214)$$

When the line is reasonably well terminated, the reflection coefficients are small. Approximate expressions may then be used to represent the logarithms of factors involving these coefficients if we apply the expansion

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$$

for the region of small values of x . Thus we find

$$\left. \begin{aligned} \gamma' &\cong \alpha l + r_S r_R (1 - e^{-2\alpha l}) \\ \gamma_E &\cong \alpha l - r_R (1 - e^{-2\alpha l}) \\ \gamma_I &\cong \alpha l + r_R (1 - e^{-2\alpha l}) \\ \gamma_{EI} &\cong \alpha l + \frac{1}{2} r_R^2 (1 - e^{-4\alpha l}) \end{aligned} \right\} \quad (214a)$$

For the first three of these only one term in the expansion is used; for the last, two terms are necessary since the odd terms contribute nothing. It should be noticed that in γ_E and γ_I the reflection is a *first order effect*, whereas in γ' and γ_{EI} it is a *second order effect*. This means that poor terminal conditions affect the quality of voltage and current transmission individually to a greater extent than they do either the volt-ampere transmission ratio or any of the insertion ratios.

The departure of the functions α and Z_0 from their ideal behavior gives rise to departures in the effective propagation function γ from its ideal behavior for the various cases summarized in (214a). If we denote the average values of the attenuation and phase functions throughout given frequency ranges by

$$\left. \begin{aligned} \alpha_{1av} &= \sqrt{RG} \cdot f_{1av} \\ \alpha_{2av} &= \omega \sqrt{LC} \cdot f_{2av} \end{aligned} \right\},$$

where f_{1av} and f_{2av} are the average values of the functions $f_1(x)$ and $f_2(x)$ discussed in section 3, then in the expression

$$\gamma = \alpha_{1av} l (1 + \delta_1) + j \alpha_{2av} l (1 + \delta_2) \quad (214b)$$

δ_1 and δ_2 represent the *net errors* (or decimal departures from the ideal) in the effective propagation function γ . The exact determination of δ_1 and δ_2 for the general case involves nothing fundamentally new but leads to expressions too cumbersome to reproduce here. We shall confine ourselves to an approximate treatment of the case for which the attenuation or line length are sufficiently great so that the factors $e^{-2\alpha l}$ or $e^{-4\alpha l}$ in (214a) may be neglected. The results of such a study may still be used as estimates in borderline cases.

For the propagation function α we have the representation

$$\alpha = \alpha_{1av} (1 + \delta_{\alpha 1}) + j \alpha_{2av} (1 + \delta_{\alpha 2}), \quad (215)$$

where $\delta_{\alpha 1}$ and $\delta_{\alpha 2}$ are the maximum decimal tolerances discussed in section 5. We recall that they are equal in magnitude but opposite in sign. Equation (180) gives $\delta_{\alpha 1}$ as

$$\delta_{\alpha 1} = -\delta_{\alpha 2} = \frac{(a - b)^2}{16 \omega_1^2},$$

where ω_1 is the lowest essential frequency.

In order to get an average maximum value for the reflection coefficients, we write for the characteristic impedance

$$\begin{aligned} Z_0 &= |Z_0|_{av} \cdot (1 + \delta_{z_0}) \cdot |\delta_z| \\ &\cong |Z_0|_{av} \cdot (1 + \delta_{z_0}) (1 + j \delta_z) \\ &\cong |Z_0|_{av} \cdot (1 + \delta_{z_0} + j \delta_z) \end{aligned}$$

where $|Z_0|_{av}$ is its average magnitude in the prescribed range. If the terminal impedance Z_R is made equal to this average value, the reflection coefficient becomes

$$r_R \cong -\frac{\delta_{z_0} + j \delta_z}{2 + \delta_{z_0} + j \delta_z} \cong -\frac{1}{2} (\delta_{z_0} + j \delta_z). \quad (215a)$$

ETHERFORCE

Assuming that the line is similarly terminated at the sending end, the coefficient r_s has the same value. Thus, substituting (215) and (215a) into (214a), and dropping the exponential terms, the net errors according to (214b) are for the function γ'

$$\delta_1 = \delta_{\alpha 1} + \frac{\delta_{z_0}^2 - \delta_{\zeta}^2}{4\alpha_{1av}l}; \delta_2 = \delta_{\alpha 2} + \frac{\delta_{z_0}\delta_{\zeta}}{2\alpha_{2av}l}; \quad (216)$$

for the functions γ_E and γ_I (top and bottom signs respectively)

$$\delta_1 = \delta_{\alpha 1} \pm \frac{\delta_{z_0}}{2\alpha_{1av}l}; \delta_2 = \delta_{\alpha 2} \pm \frac{\delta_{\zeta}}{2\alpha_{2av}l}; \quad (216a)$$

and for the function γ_{EI}

$$\delta_1 = \delta_{\alpha 1} + \frac{\delta_{z_0}^2 - \delta_{\zeta}^2}{8\alpha_{1av}l}; \delta_2 = \delta_{\alpha 2} + \frac{\delta_{z_0}\delta_{\zeta}}{4\alpha_{2av}l}. \quad (216b)$$

These results show that for long lines and high attenuation the net errors are substantially given by $\delta_{\alpha 1}$ and $\delta_{\alpha 2}$ in every case. In order to evaluate them still further, we shall assume that a is large compared to b , which is justified even though the line parameters are adjusted for reasonably small values of $\delta_{\alpha 1}$ and $\delta_{\alpha 2}$, and then have

$$\left. \begin{aligned} \delta_{\alpha 1} &= -\delta_{\alpha 2} = \frac{a^2}{16\omega_1^2} \\ \delta_{z_0} &= -\frac{a^2}{8\omega_1^2} = -2\delta_{\alpha 1} \\ \delta_{\zeta} &= -\delta_{\alpha 1}^{1/2} \left(1 - \frac{16}{3}\delta_{\alpha 1} \right) \end{aligned} \right\},$$

where (212) is used for δ_{z_0} , and an additional term in (207b) is calculated for the determination of δ_{ζ} according to (213). This is necessary in order to evaluate (216) and (216b) which involve the difference of $\delta_{z_0}^2$ and δ_{ζ}^2 . The second term in δ_{ζ} , therefore, contributes a correction of the same order of magnitude as $\delta_{z_0}^2$.

For α_{1av} and α_{2av} we can use

$$\alpha_{1av} \cong \sqrt{RG} m \cong \frac{R}{2} \sqrt{\frac{C}{L}}$$

$$\alpha_{2av} \cong \omega_1 \sqrt{LC} = \frac{1}{2} \delta_{\alpha 1}^{-1/2} \left(\frac{R}{2} \sqrt{\frac{C}{L}} \right)$$

where the frequency limit ω_1 is substituted in order to obtain the maximum error. With these values we find for γ' the errors

$$\delta_1 = \delta_{\alpha 1} \left\{ 1 + \left(\frac{11}{3}\delta_{\alpha 1} - \frac{1}{4} \right) \left(\frac{R}{2} \sqrt{\frac{C}{L}} \right)^{-1} \right\}; \delta_2 = \delta_{\alpha 2} \left\{ 1 + 2\delta_{\alpha 2} \left(\frac{R}{2} \sqrt{\frac{C}{L}} \right)^{-1} \right\}, \quad (216c)$$

for γ_E and γ_I

$$\delta_1 = \delta_{\alpha 1} \left\{ 1 \mp \left(\frac{R}{2} \sqrt{\frac{C}{L}} \right)^{-1} \right\}; \delta_2 = \delta_{\alpha 2} \left\{ 1 \pm \left(\frac{R}{2} \sqrt{\frac{C}{L}} \right)^{-1} \right\}, \quad (216d)$$

and for γ_{EI}

$$\delta_1 = \delta_{\alpha 1} \left\{ 1 + \left(\frac{11}{6}\delta_{\alpha 1} - \frac{1}{8} \right) \left(\frac{R}{2} \sqrt{\frac{C}{L}} \right)^{-1} \right\}; \delta_2 = \delta_{\alpha 2} \left\{ 1 + \delta_{\alpha 2} \left(\frac{R}{2} \sqrt{\frac{C}{L}} \right)^{-1} \right\}. \quad (216e)$$

As an illustration of this result consider the numerical example given at the end of section 5, page 98, where the necessary R/L ratio to make $\delta_{\alpha} = 0.05$ for $\omega_1 = 2\pi \cdot 100$ was calculated for the open line data on page 90. Assuming that the resistance remains unchanged, this calls for a new inductance parameter

$$L_{\text{new}} = 3.78 \times 0.004 = 0.01512.$$

With this value

$$\frac{R}{2} \sqrt{\frac{C}{L}} = 5 \sqrt{\frac{0.008}{0.01512}} \cdot 10^{-3} = 0.00364.$$

This line will have to be several hundred miles long before the reflection effects due to imperfect terminations become negligible. Suppose the line is 275 miles long. Then

$$e^{-2\alpha_1 l} = 0.135; e^{-4\alpha_1 l} = 0.018$$

so that there is little error in dropping the exponential terms in (214a).

Using $\delta_{\alpha 1} = -\delta_{\alpha 2} = 0.05$ we have for γ'

$$\delta_1 = 0.05 (1 - 0.23) = 3.8\%$$

$$\delta_2 = -0.05 (1 - 0.1) = -4.5\%$$

for γ_E

$$\delta_1 = 0.05 (1 - 1) = 0\%$$

$$\delta_2 = -0.05 (1 + 1) = 10\%$$

for γ_I

$$\delta_1 = 0.05 (1 + 1) = 10\%$$

$$\delta_2 = -0.05 (1 - 1) = 0\%$$

and for γ_{EI}

$$\delta_1 = 0.05 (1 - 0.11) = 4.5\%$$

$$\delta_2 = -0.05 (1 - 0.05) = 4.8\%$$

These are the approximate net amounts of amplitude and phase distortion over this circuit at the lowest frequency. At the highest frequency the tolerances $\delta_{\alpha 1}$ and $\delta_{\alpha 2}$ have opposite signs while δ_{ζ} is practically zero. The errors in γ' and γ_{EI} will not change appreciably. Those in γ_E and γ_I will change by about the same amount owing to a second order

effect. It is interesting to note that sometimes the reflection effects tend to cancel the errors due to the propagation function.

From this discussion we see that reflections have little effect upon the quality of transmission only when the line is long or the attenuation high. It may, of course, be that in a given situation the degree of mismatch at the line terminals is an even more important factor than the quality of transmission. This will be true at any terminal where line balance is an important item, as for example at repeater points in a two-wire single channel carrier system. There the repeater gain is definitely limited by the accuracy of balance which can be obtained between the characteristic impedance of the line and the balancing network. The situation can obviously be met by either improving the characteristics of the line sufficiently or by designing a more elaborate line balancing network. The latter process, the theory of which we shall discuss in a later chapter, is the one favored in recent practice.

9. Approximations in the consideration of the attenuation and phase functions. It often occurs that the line parameters in an actual case have such values that within certain frequency ranges one or more may be neglected without having the resulting approximate expressions for α_1 and α_2 be in error by more than a tolerable amount. When this is so, the treatment of the transmission problem may be considerably simplified numerically. It is, therefore, of value to establish criteria whereby we shall be able to predetermine whether or not such approximations are possible when the essential frequency range and the maximum allowable error are specified. This we propose to do now.

Consider the following analytic forms for α_1 and α_2 which follow from (146b), (147b), (150a), and (151a) with the use of (148), namely

$$\left. \begin{aligned} \alpha_1 &= \sqrt{RGx} (\sqrt{m^2 + y^2} - y)^{1/2} \\ \alpha_2 &= \sqrt{RGx} (\sqrt{m^2 + y^2} + y)^{1/2} \end{aligned} \right\}. \quad (217)$$

According to (153) the frequency region for which x lies in the vicinity of unity is that for which y lies in the vicinity of zero. For this vicinity the right-hand factors in (217) may be expanded as follows

$$(\sqrt{m^2 + y^2} \pm y)^{1/2} = \sqrt{m} \left\{ 1 \pm \frac{1}{2} \left(\frac{y}{m} \right) + \frac{1}{8} \left(\frac{y}{m} \right)^2 \pm \dots \right\}. \quad (218)$$

Then if we recognize that

$$\begin{aligned} \sqrt{RGmx} &= \sqrt{\frac{RC\omega}{2}} \left(1 + \frac{b}{a} \right)^{1/2} \\ &= \sqrt{\frac{RC\omega}{2}} \left\{ 1 + \frac{1}{2} \left(\frac{b}{a} \right) - \frac{1}{8} \left(\frac{b}{a} \right)^2 + \dots \right\}, \end{aligned} \quad (219)$$

and assume

$$\frac{b}{a} \ll 1 \quad (220)$$

so that to a first approximation the series in (218) and (219) may be broken off after the second terms, we can write

$$\left. \begin{aligned} \alpha_1 &\cong \sqrt{\frac{RC\omega}{2}} \left\{ 1 + \frac{1}{2} \left(\frac{b}{a} - \frac{y}{m} \right) \right\} \\ \alpha_2 &\cong \sqrt{\frac{RC\omega}{2}} \left\{ 1 + \frac{1}{2} \left(\frac{b}{a} + \frac{y}{m} \right) \right\} \end{aligned} \right\}. \quad (221)$$

But, using (148) and (153) again, we see that

$$\frac{y}{m} = \left(\frac{\omega}{a} - \frac{b}{\omega} \right) \left(1 + \frac{b}{a} \right)^{-1}, \quad (222)$$

and expanding the right-hand factor in this we have

$$\left(\frac{b}{a} \pm \frac{y}{m} \right) = \left(\frac{b}{a} \right) \pm \left(\frac{\omega}{a} - \frac{b}{\omega} \right) \mp \left(\frac{b}{a} \right) \left(\frac{\omega}{a} - \frac{b}{\omega} \right) \pm \dots$$

Retaining the first two terms in this expansion, we may write

$$\left. \begin{aligned} \alpha_1 &= \sqrt{\frac{RC\omega}{2}} (1 + \delta_1) \\ \alpha_2 &= \sqrt{\frac{RC\omega}{2}} (1 + \delta_2) \end{aligned} \right\}, \quad (223)$$

and for the decimal errors δ_1 and δ_2 have

$$\left. \begin{aligned} \delta_1 &\cong \frac{1}{2} \left(\frac{b}{a} - \frac{\omega}{a} + \frac{b}{\omega} \right) \\ \delta_2 &\cong \frac{1}{2} \left(\frac{b}{a} + \frac{\omega}{a} - \frac{b}{\omega} \right) \end{aligned} \right\}. \quad (224)$$

These results state that, if $(b/a) \ll 1$, we may use for α_1 and α_2 the approximate expression

$$\alpha_1 = \alpha_2 = \sqrt{\frac{RC\omega}{2}} \quad (225)$$

where the errors and their dependence upon frequency are approximately expressed by (224). Recalling the original form (142) for α , we recognize that the form (225) is the result of neglecting the line parameters L and G . In practice this is frequently done in the treatment of cable problems, where it is pointed out that, owing to the close proximity of the conductors and the relatively good insulation, the inductance

ATHERFORCE

and leakage are so small as to be altogether negligible in comparison with the resistance and capacitance parameters. This may very well be so, but it is nevertheless essential to demonstrate the correctness of such a procedure, particularly in doubtful cases. Furthermore it is all-important to note that whenever we use the term "negligible" it is implied that, although an error is involved, it is for the case at hand so small that it does not affect the result to a greater degree than can be tolerated. Thus the question as to whether or not certain parameters may be neglected is not adequately met unless we are able to specify the maximum magnitude of the error which will be involved when this approximation is made over a specified frequency range. The latter is

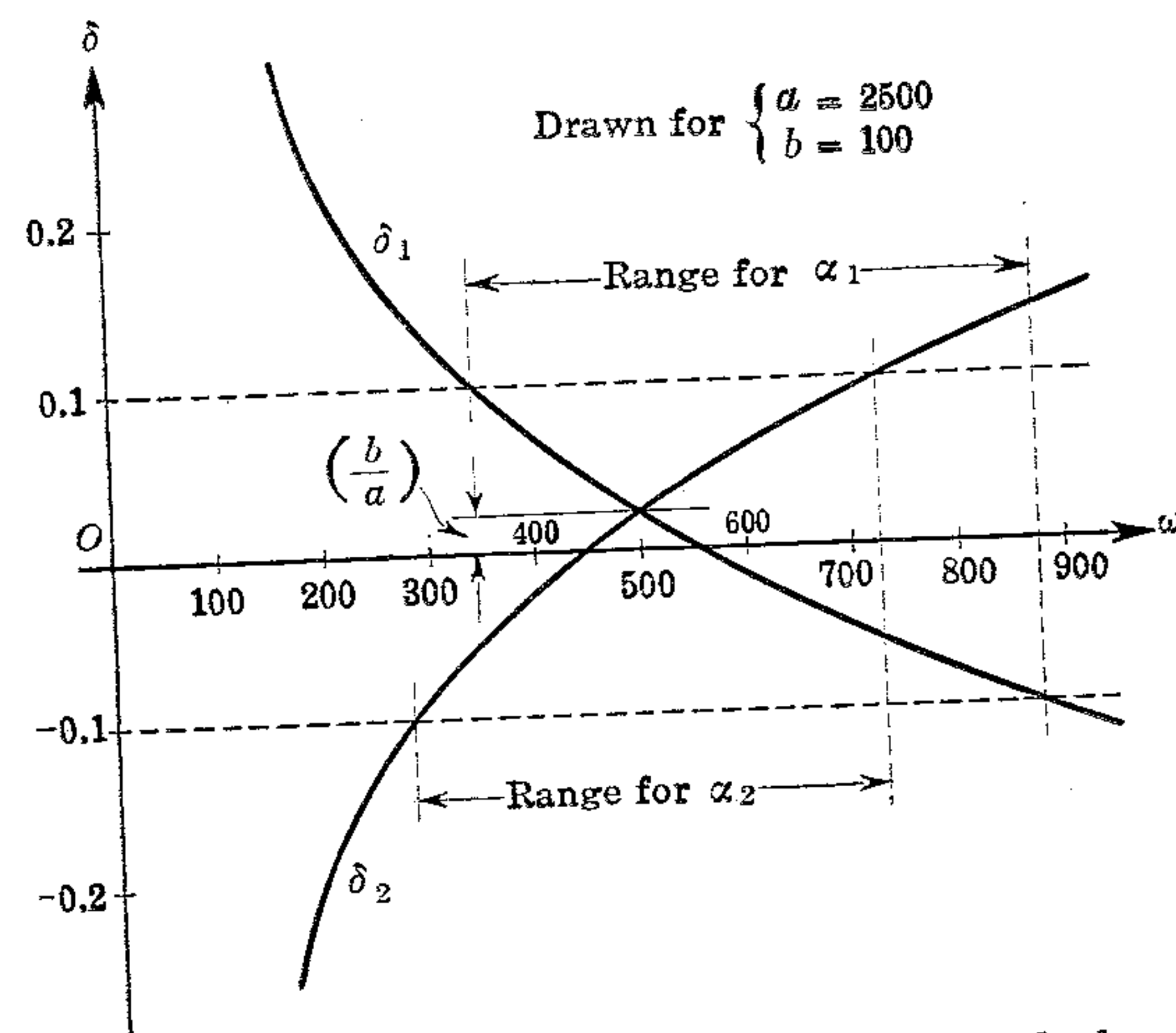


FIG. 27.—Errors in the attenuation (δ_1) and phase (δ_2) functions due to neglecting the parameters L and G showing their variations with frequency. Drawn for the data (230).

these curves we see that the errors are zero at one frequency, and increase in magnitude as this frequency is departed from. Hence, if we specify a given maximum error (in the figure this is chosen as $\delta = \pm 0.1$) we will find corresponding frequency ranges within which this error is not exceeded. We note from the figure that these ranges are not quite the same for α_1 and α_2 . This is due to the term (b/a) in (224). The smaller (b/a) , the more nearly identical will the ranges for α_1 and α_2 be. It is also interesting to note that the error has opposite signs at the two frequency limits.

When the maximum tolerance is specified and it is desired to determine the corresponding frequency ranges, we must solve the relations

a particularly significant factor in this connection, for although it may be quite adequate to neglect L and G in a cable over a certain frequency range, the same may not be true for another range unless the maximum tolerance is also modified.

This situation is expressed by the relations (224), the significance of which is most easily understood by referring to Fig. 27 where these errors are plotted versus frequency for a numerical example. From

(224) for ω . This gives for the ranges for α_1 and α_2 respectively

$$\frac{-(2a\delta_1 - b) + \sqrt{(2a\delta_1 - b)^2 + 4ab}}{2} \leq \omega \leq \frac{(2a\delta_1 + b) + \sqrt{(2a\delta_1 + b)^2 + 4ab}}{2} \quad (226)$$

and

$$\frac{-(2a\delta_2 + b) + \sqrt{(2a\delta_2 + b)^2 + 4ab}}{2} \leq \omega \leq \frac{(2a\delta_2 - b) + \sqrt{(2a\delta_2 - b)^2 + 4ab}}{2} \quad (227)$$

where the δ 's represent the magnitudes of the errors only. If

$$b \ll 2a\delta \quad (228)$$

these expressions may be approximated by neglecting b in comparison with $2a\delta$, giving the following range common to α_1 and α_2

$$a\delta \left\{ \sqrt{1 + \frac{b}{a\delta^2}} - 1 \right\} \leq \omega \leq a\delta \left\{ \sqrt{1 + \frac{b}{a\delta^2}} + 1 \right\} \quad (229)$$

This may be further simplified if

$$b \ll a\delta^2 \quad (228a)$$

in which case the radicals may be expanded and only the first two terms retained. Then the common range is expressed by

$$\frac{b}{2\delta} \left(1 - \frac{b}{4a\delta^2} \right) \leq \omega \leq 2a\delta \left(1 + \frac{b}{4a\delta^2} \right) \quad (229a)$$

Note that where the use of (229) or (229a) is justified the product of the limiting frequencies equals ab ; i.e., the point $x = 1$ becomes the geometric mean between the limits of the range. This shows that the vicinity of $x = 1$ is that region within which the inductance and leakage are negligible provided to begin with

$$\left(\frac{b}{a} \right) = \frac{LG}{RC} \ll 1.$$

In order to illustrate these results consider first the data for an open telephone line

$$\left. \begin{aligned} R &= 10; L = 0.004 \\ G &= 0.8 \cdot 10^{-6}; C = 0.008 \cdot 10^{-6} \end{aligned} \right\} \quad (230)$$

Here

$$a = \frac{R}{L} = 2500; b = \frac{G}{C} = 100; \frac{b}{a} = \frac{1}{25}.$$

Suppose we let

$$\delta_1 = \delta_2 = 0.1,$$

i.e., fix the maximum error at 10 per cent. Although

$$\frac{b}{2a\delta} = \frac{1}{5}$$

we shall not consider (228) sufficiently justified. Hence we use the forms (226) and (227) and find for α_1 the range

$$\left. \begin{aligned} 339 &\leq \omega \leq 883 \\ 54 &\leq f \leq 140 \end{aligned} \right\}, \quad (231)$$

or

and for α_2 the range

$$\left. \begin{aligned} 283 &\leq \omega \leq 739 \\ 45 &\leq f \leq 118 \end{aligned} \right\}. \quad (232)$$

or

These results will be seen to check the Fig. 27 which is drawn for this case. The frequency ranges thus found are not very extensive, to be sure, but it is rather surprising that both the inductance and the leakage are at all negligible on an ordinary open telephone line, even to within a maximum error of 10 per cent and over a limited frequency range.

In contrast to this, consider the following telephone cable data

$$\left. \begin{aligned} R &= 171; L = 0.001 \\ G &= 1.75 \cdot 10^{-6}; C = 0.073 \cdot 10^{-6} \end{aligned} \right\}. \quad (233)$$

Then

$$a = \frac{R}{L} = 171,000; b = \frac{G}{C} = 24; \frac{b}{a} = \frac{1}{7125}.$$

Again let

$$\delta_1 = \delta_2 = 0.1.$$

Here

$$\frac{b}{a\delta^2} = \frac{1}{71},$$

so that the condition (228a) is sufficiently satisfied, and hence we can use the relation (229a) for the determination of the common frequency range. We find

$$\left. \begin{aligned} 140 &\leq \omega \leq 34,200 \\ 22 &\leq f \leq 5450. \end{aligned} \right\} \quad (234)$$

or

Thus we see that on this cable circuit the inductance and leakage are negligible, so far as the propagation function is concerned, to within a maximum error of 10 per cent over a frequency range of 22 to 5450

cycles per second. This is a considerable range, and adequately covers the voice spectrum, but it is nevertheless significant that below 22 and above 5450 cycles per second even this cable circuit cannot be treated by considering its resistance and capacitance alone without introducing an error in excess of 10 per cent.

If the conditions or the requirements are such that for a given circuit the above approximations cannot be made, then it may still be possible that for this circuit the leakage parameter alone is negligible. In order to study this possibility, consider again the forms (217), but in the slightly modified form

$$\left. \begin{aligned} \alpha_1 &= \sqrt{RGmx} \left(\sqrt{1 + \left(\frac{y}{m}\right)^2} - \frac{y}{m} \right)^{1/2} \\ \alpha_2 &= \sqrt{RGmx} \left(\sqrt{1 + \left(\frac{y}{m}\right)^2} + \frac{y}{m} \right)^{1/2} \end{aligned} \right\}. \quad (235)$$

Now let

$$\left. \begin{aligned} \frac{y}{m} &= y_a + y_b \\ y_a &= \frac{\omega}{a} \end{aligned} \right\}, \quad (236)$$

and

so that by (222)

$$y_b = - \left(\frac{b}{\omega} + \frac{b}{a} \frac{\omega}{a} \right) \left(1 + \frac{b}{a} \right)^{-1}. \quad (237)$$

Since by using (148) we may modify (219) so as to give either

$$\sqrt{RGmx} = \frac{R}{2} \sqrt{\frac{C}{L}} (2y_a)^{1/2} \left(1 + \frac{b}{a} \right)^{1/2} \quad (238)$$

or

$$\sqrt{RGmx} = \omega \sqrt{LC} (2y_a)^{-1/2} \left(1 + \frac{b}{a} \right)^{1/2}, \quad (239)$$

we recognize that if we can determine the frequency range for which y_b is negligible, we will also have the region of negligible G . This is so because y_b vanishes for $b = 0$, which is the same as $G = 0$. For this condition (238) and (239) merely drop their last factor.

In order to determine this range and the corresponding errors in α_1 and α_2 , we assume

$$y_b \ll y_a$$

and expand the second factors in (235) about the point $y_b = 0$, retaining only the first two terms. This gives

$$\left(\sqrt{1 + \left(\frac{y}{m}\right)^2} \pm \frac{y}{m} \right)^{1/2} = (\sqrt{1 + y_a^2} \pm y_a)^{1/2} \left(1 \pm \frac{y_b}{2\sqrt{1 + y_a^2}} \right). \quad (240)$$

ETHERFORCE

Substituting this together with (237), (238), and (239) into (235), and using the first two terms of

$$\left(1 + \frac{b}{a}\right)^{1/2} = 1 + \frac{b}{2a} - \frac{1}{8}\left(\frac{b}{a}\right)^2 + \dots,$$

we find

$$\left. \begin{aligned} \alpha_1 &= \frac{R}{2} \sqrt{\frac{C}{L}} (2y_a)^{1/2} (\sqrt{1 + y_a^2} - y_a)^{1/2} (1 + \delta_1) \\ \alpha_2 &= \omega \sqrt{LC} (2y_a)^{-1/2} (\sqrt{1 + y_a^2} + y_a)^{1/2} (1 + \delta_2) \end{aligned} \right\} \quad (241)$$

and

with

$$\delta_{1,2} = \frac{b}{2a} \pm \frac{\left(\frac{b}{\omega} + \frac{b}{a} \frac{\omega}{a}\right)}{2\left(1 + \frac{b}{a}\right) \sqrt{1 + \left(\frac{\omega}{a}\right)^2}}. \quad (242)$$

The forms (241) are the result of neglecting G , and (242) determines the errors and their frequency dependence. A study of (242) shows that the errors are infinite at zero frequency, and decrease uniformly with increasing frequency approaching an asymptotic value as the frequency tends toward infinity. More specifically

$$\left. \begin{aligned} &\text{for } \omega \rightarrow \infty \\ &\delta_1 \rightarrow \frac{b}{a} \\ &\delta_2 \rightarrow \frac{1}{2} \left(\frac{b}{a}\right)^2 \end{aligned} \right\}, \quad (243)$$

and to within a sufficient approximation

$$\left. \begin{aligned} &\text{for } \omega \rightarrow 0 \\ &\delta_1 \rightarrow \left(\frac{b}{2a} + \frac{b}{2\omega}\right) \\ &\delta_2 \rightarrow \left(\frac{b}{2a} - \frac{b}{2\omega}\right) \end{aligned} \right\}. \quad (244)$$

Since it is assumed to start with that $b \ll a$, the errors (243) are negligible. Hence if the asymptotic values (243) are sufficiently small for the case at hand, the approximate relations (244) may be used to determine the lower frequency limit below which a prescribed error will be exceeded. This is easily done by solving (244) for ω . Thus the range for negligible G is given by

$$\left. \begin{aligned} \frac{ab}{2a\delta_1 - b} &\leq \omega \leq \infty \\ \frac{ab}{2a\delta_2 + b} &\leq \omega \leq \infty \end{aligned} \right\}, \quad (245)$$

or if

$$\frac{b}{2a\delta} \ll 1, \quad (246)$$

these are with sufficient accuracy

$$\left. \begin{aligned} \frac{b}{2\delta_1} \left(1 + \frac{b}{2a\delta_1}\right) &\leq \omega \leq \infty \\ \frac{b}{2\delta_2} \left(1 - \frac{b}{2a\delta_2}\right) &\leq \omega \leq \infty \end{aligned} \right\}. \quad (245a)$$

Applying these results to the data (230) for an average open line we find that the leakage parameter is negligible to within a maximum error of 10 per cent in α_1 for

$$\left. \begin{aligned} 600 &\leq \omega \leq \infty \\ 95 &\leq f \leq \infty \end{aligned} \right\},$$

and in α_2 for

$$\left. \begin{aligned} 400 &\leq \omega \leq \infty \\ 64 &\leq f \leq \infty \end{aligned} \right\}.$$

This is quite remarkable for an ordinary open telephone line. Of course, as pointed out in Chapter I, the leakage parameter increases considerably with frequency.

The value in (230) is for a frequency of about 800 cycles. Hence, although the variation is not considered in the above calculations, it is safe to say that over the voice range, and perhaps considerably beyond, the leakage parameter is still negligible.

When the variation with frequency is known, the various forms discussed

above may obviously still be applied by using a trial method. Thus the considerable variation of the leakage parameter with frequency is found for many practical cases to cause no annoyance whatsoever.

This situation is even more true in a cable. Using the data (233) we find by (245a) that the leakage parameter is negligible in both α_1 and α_2 to within a maximum error of 10 per cent for all frequencies above 22 cycles per second.

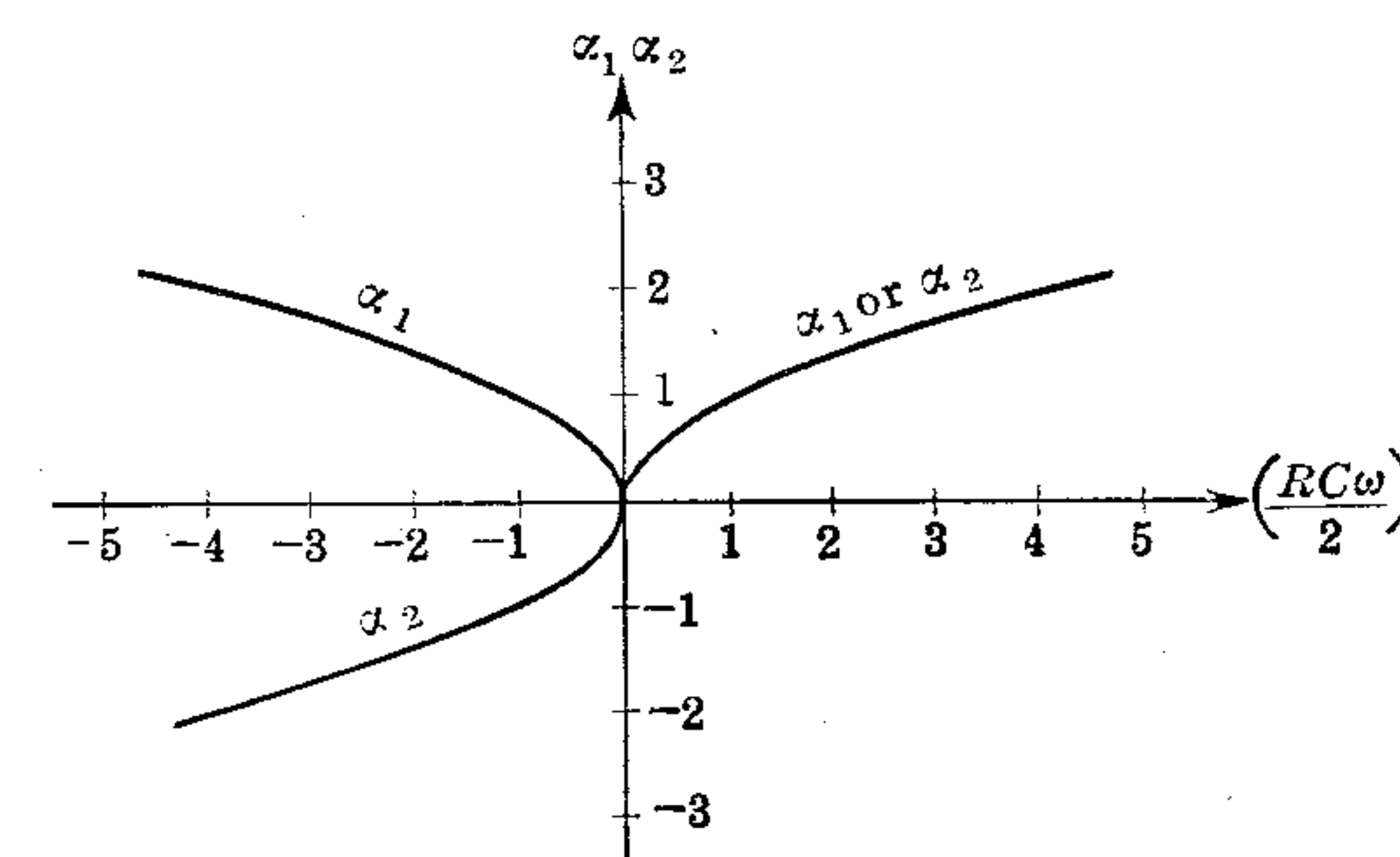


FIG. 28.—Attenuation and phase as functions of frequency when L and G are neglected and R and C assumed constant.

Thus we see that, for certain frequency ranges and tolerances, either G alone or both L and G may be neglected in the analysis of a given

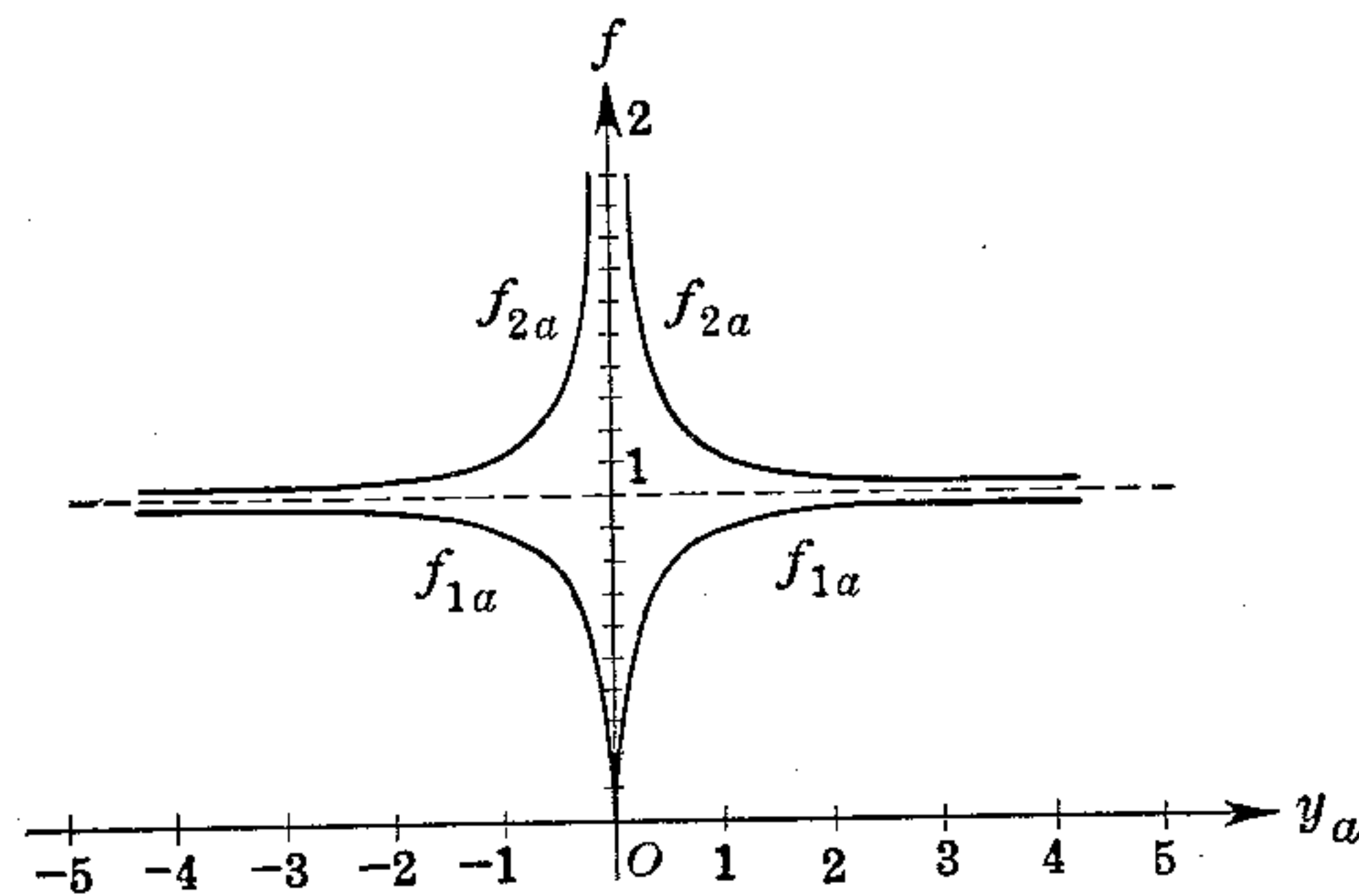


FIG. 29.—Departure functions for the attenuation and phase versus the frequency variable y_a , eq. (236), when the line parameter G is neglected.

represented by the upper, while α_2 is represented by the lower branch of the parabola. This must be so because α_1 is an even and α_2 an odd function.

When only G is negligible the simplification is not so great. We then have the expressions (241) with the δ 's dropped. If we write

$$\left. \begin{aligned} \alpha_1 &= \frac{R}{2} \sqrt{\frac{C}{L}} \cdot f_{1a} \\ \alpha_2 &= \omega \sqrt{LC} \cdot f_{2a} \end{aligned} \right\}, \quad (247)$$

then

$$\left. \begin{aligned} f_{1a} &= (2y_a)^{1/2} (\sqrt{1 + y_a^2} - y_a)^{1/2} \\ f_{2a} &= (2y_a)^{-1/2} (\sqrt{1 + y_a^2} + y_a)^{1/2} \end{aligned} \right\}. \quad (248)$$

Plots of f_{1a} and f_{2a} are shown in Fig. 29. Note that

$$f_{1a} f_{2a} = 1, \quad (249)$$

i.e., the functions are reciprocals of each other. This simplifies the plotting considerably. It should be pointed out here that the functions (248) are not even functions of y_a . In fact for negative values of y_a they become imaginary. This rather puzzling point is cleared up if we consider (247) and form $\alpha_1 + j\alpha_2$. Then for a negative y_a the first term becomes imaginary and the second real, i.e., α_1 and α_2 interchange places but their analytic forms are left unchanged. The functions in Fig. 29 for negative y_a -values are those which are obtained after this con-

facility so far as the propagation function is concerned. When this is so the propagation function and consequently the numerical work will be simplified. For example, when both L and G are negligible, we have the one simple function (225) as a representation for both α_1 and α_2 . A plot of this is shown in Fig. 28. It is simply a rectangular parabola tipped over on its side. Note that for negative frequencies α_1 is

version process has been carried out. They are, of course, images of the corresponding functions for positive arguments.

10. Approximations in the consideration of the characteristic impedance function. When the inductance and leakage parameters are negligible in the determination of the characteristic impedance function, then a simplification in its analytic form is also possible. The maximum error and corresponding frequency range will, however, be different from those for the propagation function.

Considering the magnitude function (199), this may be put into the form

$$z_0 = \left(\frac{a}{\omega}\right)^{1/2} \left(1 + \frac{\omega^2}{a^2}\right)^{1/4} \left(1 + \frac{b^2}{\omega^2}\right)^{-1/4}, \quad (250)$$

so that by (198a) we have for the magnitude of the characteristic impedance

$$|Z_0| = \sqrt{\frac{R}{C\omega}} \left(1 + \frac{\omega^2}{a^2}\right)^{1/4} \left(1 + \frac{b^2}{\omega^2}\right)^{-1/4}. \quad (251)$$

Using the first two terms of the expansions

$$\left. \begin{aligned} \left(1 + \frac{\omega^2}{a^2}\right)^{1/4} &= 1 + \frac{1}{4} \frac{\omega^2}{a^2} + \dots \\ \left(1 + \frac{b^2}{\omega^2}\right)^{-1/4} &= 1 - \frac{1}{4} \frac{b^2}{\omega^2} + \dots \end{aligned} \right\} \quad (252)$$

and writing

$$|Z_0| = \sqrt{\frac{R}{C\omega}} (1 + \delta), \quad (251a)$$

we have

$$\delta = \frac{1}{4} \left(\frac{\omega^2}{a^2} - \frac{b^2}{\omega^2} \right) \quad (253)$$

for the error and its dependence upon frequency.

The form (251a) without the factor $(1 + \delta)$ is obviously the resulting expression for the magnitude of the characteristic impedance for $L = G = 0$. From (253) we find that the frequency region within which this approximation is valid to within a maximum decimal error δ is defined by

$$\sqrt{(4a^4\delta^2 + a^2b^2)^{1/2} - 2a^2\delta} \leq \omega \leq \sqrt{(4a^4\delta^2 + a^2b^2)^{1/2} + 2a^2\delta}. \quad (254)$$

When $b \ll 2a\delta$, this range is approximately given by

$$\frac{b}{2\sqrt{\delta}} \left(1 - \frac{b^2}{32a^2\delta^2}\right) \leq \omega \leq 2a\sqrt{\delta} \left(1 + \frac{b^2}{32a^2\delta^2}\right). \quad (254a)$$

Note that the product of these frequency limits equals ab . Thus the point $x = 1$ ($\omega = \sqrt{ab}$) is the geometric mean frequency of this region.

Using the data (230) for the open line and letting $\delta = 0.1$, we find that L and G are negligible in $|Z_0|$ with a maximum error of 10 per cent for the range

$$\left. \begin{array}{l} 158 \leq \omega \leq 1580 \\ 25 \leq f \leq 250 \end{array} \right\}. \quad (255)$$

For the cable data (233), on the other hand, we find for the same maximum error the range

$$\left. \begin{array}{l} 38 \leq \omega \leq 108,000 \\ 6 \leq f \leq 17,200 \end{array} \right\}. \quad (256)$$

Let us now see whether L and G may under certain circumstances be neglected in the formulation of the angle of the characteristic impedance. From (205) we may easily get

$$\zeta(\omega) = \frac{1}{2} \left\{ \tan^{-1} \left(\frac{\omega}{a} \right) - \tan^{-1} \left(\frac{\omega}{b} \right) \right\}. \quad (257)$$

Assuming

$$\left(\frac{\omega}{a} \right) \ll 1, \text{ and } \left(\frac{\omega}{b} \right) \gg 1,$$

we may expand the inverse tangent functions and get

$$\left. \begin{array}{l} \tan^{-1} \left(\frac{\omega}{a} \right) = \left(\frac{\omega}{a} \right) - \frac{1}{3} \left(\frac{\omega}{a} \right)^3 + \dots \\ \tan^{-1} \left(\frac{\omega}{b} \right) = \frac{\pi}{2} - \left(\frac{\omega}{b} \right)^{-1} + \frac{1}{3} \left(\frac{\omega}{b} \right)^{-3} - \dots \end{array} \right\}. \quad (258)$$

If we retain only the linear terms we have for (257)

$$\zeta(\omega) = -\frac{\pi}{4} + \frac{1}{2} \left(\frac{\omega}{a} + \frac{b}{\omega} \right),$$

or if we write

$$\zeta(\omega) = -\frac{\pi}{4} (1 + \delta) \quad (259)$$

we get

$$\delta = -\frac{2}{\pi} \left(\frac{\omega}{a} + \frac{b}{\omega} \right). \quad (260)$$

The value $-\frac{\pi}{4}$ in (259) we recognize as the angle of the characteristic impedance function for neglected L and G . The relation (260) gives the error and its dependence upon frequency. Solving for the frequency range we find

$$\frac{1}{4}(\pi a \delta - \sqrt{\pi^2 a^2 \delta^2 - 16ab}) \leq \omega \leq \frac{1}{4}(\pi a \delta + \sqrt{\pi^2 a^2 \delta^2 - 16ab}), \quad (261)$$

or if $b \ll a\delta^2$ this reduces to the following approximate form

$$\frac{2b}{\pi\delta} \left(1 + \frac{4b}{\pi^2 a \delta^2} \right) \leq \omega \leq \frac{\pi a \delta}{2} \left(1 - \frac{4b}{\pi^2 a \delta^2} \right). \quad (261a)$$

Note that (261) has solutions only when

$$\left(\frac{b}{a} \right) \leq \left(\frac{\pi\delta}{4} \right)^2, \quad (262)$$

i.e., unless the ratio b/a is smaller than this value, no range exists within which L and G are negligible for the angle ζ .

To illustrate this, consider again the open line data (230), and assume a maximum error of 10 per cent. Here

$$\frac{b}{a} = 0.04$$

and

$$\left(\frac{\pi\delta}{4} \right)^2 = 0.00617.$$

Hence the criterion (262) is not satisfied, and there exists no range for which L and G may be neglected in ζ to within a maximum error of 10 per cent. Note, however, that for the functions α_1 , α_2 , and $|Z_0|$ we did find ranges within which this same error was not exceeded! The student should thus clearly recognize that, although we may be able in a given instance to neglect L and G in some functions, it does not follow that this may be generally done. This emphasizes again the point made earlier that when we say that certain parameters are negligible it is absolutely essential to mention the function for which these are neglected as well as the maximum error which we are willing to tolerate and the corresponding frequency range.

For the cable data (233) we have with respect to the angle ζ and a maximum error of 10 per cent.

$$\left(\frac{b}{a} \right) = 0.00014$$

$$\left(\frac{\pi\delta}{4} \right)^2 = 0.00617.$$

Here the criterion (262) is obviously satisfied. Also the approximate form (261a) is applicable. We thus find the range

$$\left. \begin{array}{l} 153 \leq \omega \leq 26,850 \\ 24 \leq f \leq 4,275 \end{array} \right\} \quad (263)$$

within which both L and G are negligible in ζ with a maximum error of 10 per cent for the cable whose data are given by (233).

ETHERFORCE

Finally we consider the range for which G alone may be negligible in either the magnitude or angle of the characteristic impedance function. For the magnitude we see from (250) that the leakage parameter will be negligible (this means that b is negligible) when the last factor may be replaced by unity. Using for this factor the first two terms of the expansion in (252) and writing for brevity

$$\left(\frac{\omega}{a}\right) = y_a$$

as has already been done in (236), we have

$$|Z_0| = \sqrt{\frac{L}{C}} \cdot \sqrt{\frac{1+y_a^2}{y_a^2}} (1 + \delta) \quad (264)$$

with

$$\delta = -\frac{b^2}{4\omega^2}. \quad (265)$$

This error vanishes at high frequencies. The frequency range for a given error, without regard to the sign of this error, therefore is

$$\frac{b}{2\sqrt{\delta}} \leq \omega \leq \infty. \quad (265a)$$

Note that this is practically the same as the lower frequency limit in the region for negligible L and G for $|Z_0|$ as expressed by (254a). Hence for the open line data (230), and with $\delta = 0.1$, we have from (255) that G alone is negligible in $|Z_0|$ for

$$\left. \begin{array}{l} 158 \leq \omega \leq \infty \\ 25 \leq f \leq \infty \end{array} \right\}, \quad (266)$$

and for the cable data (233) we get from (256) the range

$$\left. \begin{array}{l} 38 \leq \omega \leq \infty \\ 6 \leq f \leq \infty \end{array} \right\}. \quad (267)$$

For a possible range of negligible G in the angle ζ , consider again the form (257) and expand only the second anti-tangent according to (258), retaining the first two terms. This gives

$$\zeta \cong \frac{1}{2} \left(\tan^{-1} \left(\frac{\omega}{a} \right) - \frac{\pi}{2} + \frac{b}{\omega} \right),$$

or if we write

$$\zeta = \left(-\frac{\pi}{4} + \frac{1}{2} \tan^{-1} \frac{\omega}{a} \right) (1 + \delta) \quad (268)$$

we have

$$\delta = \frac{b}{\omega \left(\tan^{-1} \frac{\omega}{a} - \frac{\pi}{2} \right)}. \quad (269)$$

This error is shown plotted versus ω/a in Fig. 30, from which we see that with increasing frequency the asymptotic value b/a is approached, while for smaller frequencies the error tends toward infinity. Assuming, therefore, that $b < a\delta$, and that at the low-frequency end of the range

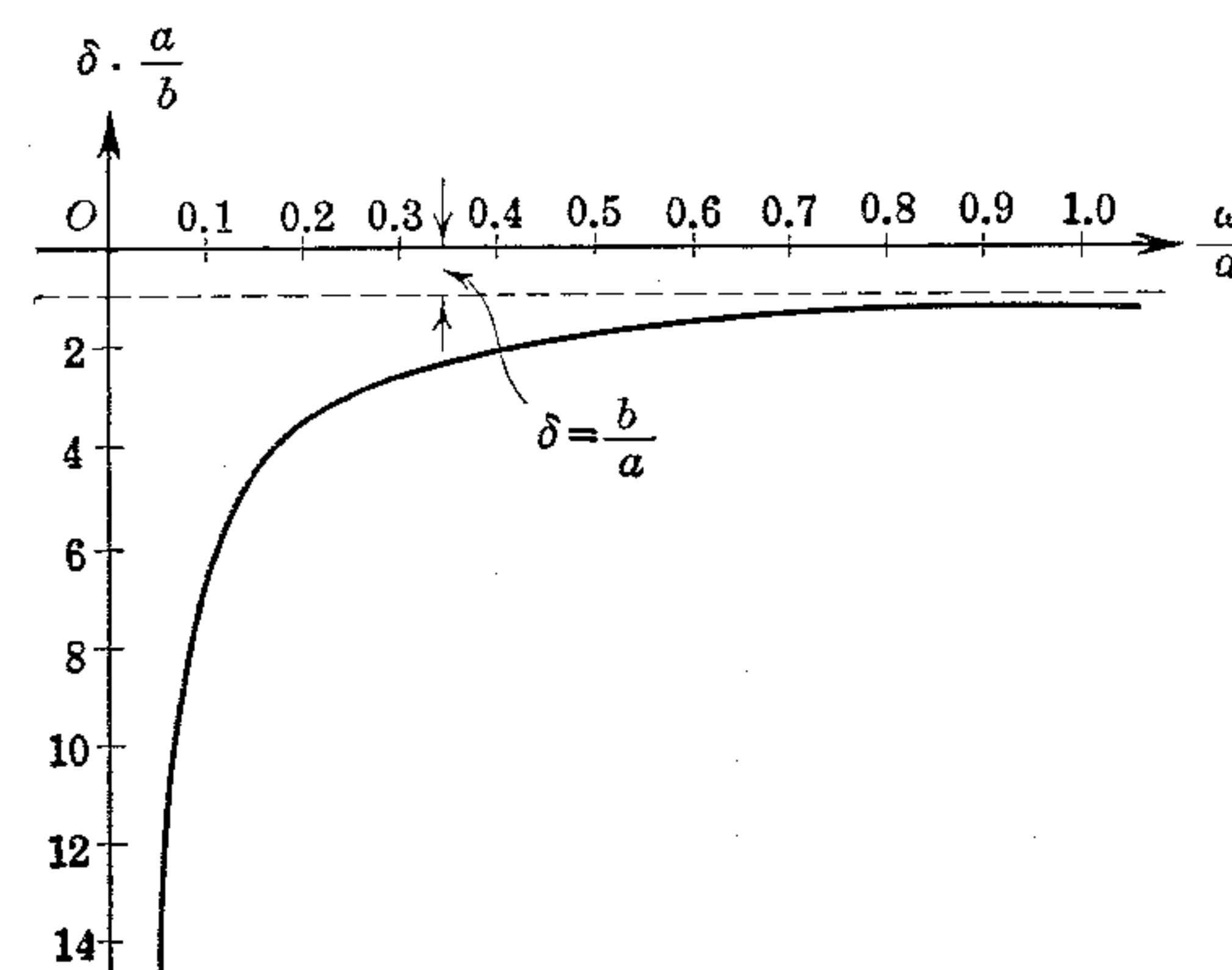


FIG. 30.—The error due to neglecting the line parameter G in the angle of the characteristic impedance function and its dependence upon frequency.

the anti-tangent in (269) is with sufficient accuracy given by its argument, we find that G is negligible in ζ to within a maximum error δ for the range

$$\frac{a\pi}{4} \left\{ 1 - \sqrt{1 - \frac{16b}{a\pi^2\delta}} \right\} \leq \omega \leq \infty, \quad (270)$$

or if $b \ll a\delta$ we may use the approximate form

$$\frac{2b}{\pi\delta} \left(1 + \frac{4b}{\pi^2 a\delta} \right) \leq \omega \leq \infty. \quad (270a)$$

For the open line data (230) this gives for $\delta = 0.1$

$$\left. \begin{array}{l} 800 \leq \omega \leq \infty \\ 127 \leq f \leq \infty \end{array} \right\}, \quad (271)$$

while for the cable data (233) the range for negligible G is

$$\left. \begin{array}{l} 153 \leq \omega \leq \infty \\ 24 \leq f \leq \infty \end{array} \right\}. \quad (272)$$

Thus under circumstances where L and G are negligible the magnitude of the characteristic impedance may be represented by the simple function (251a) without the $(1 + \delta)$ factor while the angle may be assumed constant and equal to $-\frac{\pi}{4}$ for positive frequencies. For such cases the analytic treatment may, therefore, be considerably simplified.

When circumstances permit only that G be neglected, the analysis may be somewhat simplified by using (264) without the $(1 + \delta)$ factor for $|Z_0|$, and (268) without this factor for the angle ζ . Whenever such approximations are considered, however, it is always necessary first to determine whether, for the essential frequency range involved, the maximum error remains below a value consistent with the degree of accuracy required.

PROBLEMS TO CHAPTER III

3-1. For the open-line data: $R = 10.4$; $L = 0.00367$; $C = 0.00835 \cdot 10^{-6}$; $G = 0.8 \cdot 10^{-6}$, determine and plot the departure functions $f_1(x)$ and $f_2(x)$ as well as the attenuation and phase velocity over a frequency range of 0 to 5000 cycles per second.

3-2. Repeat Problem 3-1 for the cable data: $R = 85.2$; $L = 0.001$; $C = 0.066 \cdot 10^{-6}$; $G = 1.585 \cdot 10^{-6}$.

3-3. For the open-line data of Problem 3-1 determine and plot the magnitude and angle and also the real and imaginary parts of the characteristic impedance function over the range 0-5000 cycles per second.

3-4. Repeat Problem 3-3 for the cable data of Problem 3-2.

3-5. Assuming that in the line of Problem 3-1 the conductance parameter is approximately constant and equal to $0.8 \cdot 10^{-6}$ up to 1000 cycles per second and then increases linearly with frequency at the rate of 10^{-6} mho per 1000 cycles per second, how does this affect the results in Problems 3-1 and 3-3?

3-6. A frequency group of the type shown in Fig. 21 consists of a total of 15 components, the frequency intervals corresponding to $\delta\omega = 500$. Determine and plot the corresponding wave envelope. For a mean frequency corresponding to $\omega_m = 5000$, determine the phase and envelope velocities for the open line of Problem 3-1 and the cable of Problem 3-2.

3-7. The cable whose data are specified in Problem 3-2 is to be loaded so that the maximum departure from constancy on the part of the attenuation function or departure from linearity on the part of the phase function does not exceed 8 per cent over a frequency range of 100 to 3500 cycles per second. Determine approximately the necessary R/L ratio which will meet these specifications. Assuming that the R/L ratio of the loading coils is constant and equal to 80, and that the spacing is sufficiently close so that the loading may be considered approximately uniform, determine the resulting average attenuation and compare with the maximum and minimum attenuation values at the extremes of the frequency range for the unloaded circuit.

3-8. Determine the corresponding approximate maximum departures from ideal behavior in the magnitude and angle of the characteristic impedance function of the loaded circuit of Problem 3-7, assuming the loading to be uniformly distributed; determine and plot the magnitude and angle of Z_0 and compare with the results of Problem 3-4.

3-9. The cable circuit whose data are specified in Problem 3-2 is loaded by inserting coils having a resistance of 10.6 ohms and an inductance of 0.175 henry at intervals of 1.66 miles. Assuming this amount of loading to be uniformly distributed, find the resulting maximum departures from ideal behavior on the part of the attenuation, phase, and characteristic impedance functions over the frequency ranges: $100 < f < 4000$ and $200 < f < 3000$, and compare with the de-

partures found for the corresponding non-loaded facility from the results of Problems 3-1 to 3-4. Also determine and compare the average attenuation and delay per mile of the loaded and unloaded circuits.

3-10. An open telephone line has the following parameter values: $R = 4.14$; $L = 0.00337$; $C = 0.00914 \cdot 10^{-6}$; $G = 0.8 \cdot 10^{-6}$.

(a) Determine the range of frequency within which both L and G may be neglected in the propagation function without exceeding an approximate error of 10 per cent.

(b) Determine the range of frequencies within which G alone may be neglected in the propagation function to within this same maximum error.

3-11. Consider the statement of Problem 3-10 with reference to the magnitude and angle of the characteristic impedance function.

3-12. Repeat Problem 3-10 for the cable data of Problem 3-2.

3-13. Repeat Problem 3-11 for the cable data of Problem 3-2.

3-14. Reconsider Problem 3-10, assuming the conductance parameter to vary as described in Problem 3-5, and use a cut-and-try procedure to determine how the results are affected.

CHAPTER IV

CHARACTERISTICS OF FOUR-TERMINAL NETWORKS

1. **Significance of the problem.** In the practical consideration of transmission-line problems we are seldom interested in the distribution of current or voltage along the line, i.e., in what goes on within the structure itself. Primarily the transmission problem concerns itself with the output relative to the input. This was touched upon in sections 15 and 17 of Chapter II where we developed the relations between input and output voltages and currents and discussed some of their ratios. Viewed from this angle, the transmission line or cable is essentially a network having two input and two output terminals, the

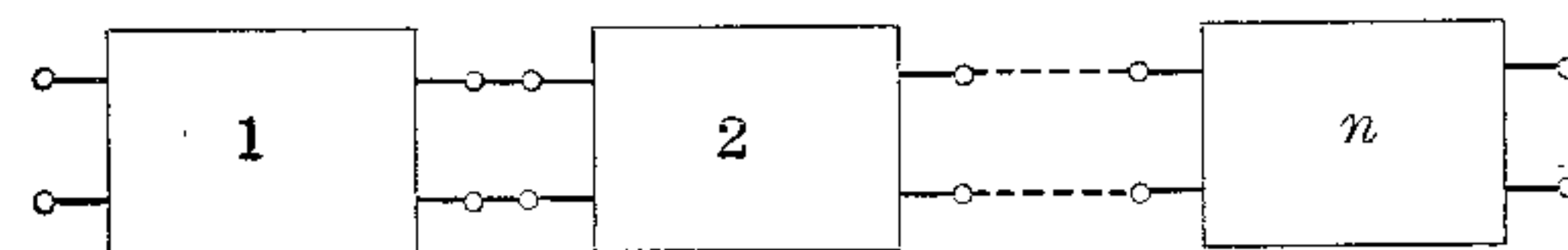


FIG. 31.—Schematic representation of a communication channel consisting of several four-terminal networks in cascade.

character of the intervening structure being immaterial except in so far as it determines the behavior of the network as a whole. From the standpoint of the analysis of its external per-

formance we could think of the transmission line as a box with four terminals, two of which are designated as input and two as output terminals. The box itself can be thought of as containing a linear network of some kind whose characteristics are such as to give rise to the external behavior already discussed. This way of viewing the transmission-line problem brings the long line into the same category with a large number of lumped-constant networks which play various important rôles in the complete transmission system as it exists in practice. Such a complete system usually consists not only of the line or cable itself but in addition contains numerous other links whose purpose it is to correct defects in the line performance as well as supply selective or other characteristics to the system which may be required by the class of service which is expected of the facility as a whole.

A single link of a complete transmission system, which may be thought of as a box with two pairs of terminals, is called a **four-terminal network**, or perhaps more accurately a **two-terminal pair**, since of the four terminals two are always designated as input and two as output terminals. A complete communication channel is made up of a number of such four-terminal networks usually connected in chain or cascade.

Fig. 31 illustrates such a channel in a schematic way. In more complicated systems we may encounter several other types of connection besides the cascade connection of Fig. 31. One or more of the boxes or links may be transmission lines or cables, i.e., distributed-constant networks, while the rest will be lumped-constant networks of various kinds. The performance of the distributed-constant four-terminal networks having been reduced to the same basis as that of the lumped-constant networks, the system as a whole becomes homogeneous and thus we are able to develop a consistent method of analysis for the determination of the overall performance.

In this chapter we wish to focus our attention entirely upon the external behavior of a four-terminal network in order to formulate general concepts as to how this external behavior may be uniquely characterized, how this characterization may be analytically expressed, and how it may be related to the contents of the box. In this connection the effect of various types of symmetry will be treated. Furthermore various types of interconnection of a number of four-terminal networks will be considered. If, in such an interconnection, two pairs of terminals are again designated as input and output, the combination may obviously be looked upon as a single four-terminal network. For example, the cascade arrangement in Fig. 31 may be considered as a single four-terminal network between the input of box 1 and the output of box n . The question arises as to how we are to obtain the resultant performance characteristics of such a system consisting of an interconnection of boxes when their individual properties are known.

Similarly a given four-terminal network may perhaps be more effectively treated if first resolved into simpler component parts each of which is again a four-terminal network. Thus not only the characterization of a given four-terminal network, but its resolution into component parts, as well as its combination with other networks, must be studied before we shall be able to consider a complete transmission system adequately. Incidentally, the results of this analysis will also enable us to treat the situation encountered when line loading is carried out by inserting inductance lumps at intervals, since such a procedure obviously breaks the line up into a cascade of four-terminal units. Problems regarding the design of artificial lines, line balances, filters, and corrective networks of various kinds all require a knowledge of the behavior of four-terminal networks for their attack. Such a study, therefore, seems appropriate at this time.

2. **Possible relationships between voltages and currents.** Fig. 32 is a sketch of a four-terminal network, indicating by means of arrows the conventions which we shall adopt relative to the reference directions of

currents and voltages at its two ends. The network itself is arbitrary except that it is linear and contains no sources of emf. The convention as to reference directions is a symmetrical one. This is adopted because it does not specify a particular pair of terminals as representing either the input or output end of the network, which fact makes it possible to generalize our analysis more easily. The two pairs of terminals are designated as 1-1' and 2-2'; these we shall also refer to more briefly as *end 1* and *end 2*.

The behavior of the network with regard to its two pairs of terminals is, therefore, expressible in terms of the four quantities I_1, E_1, I_2, E_2 . A

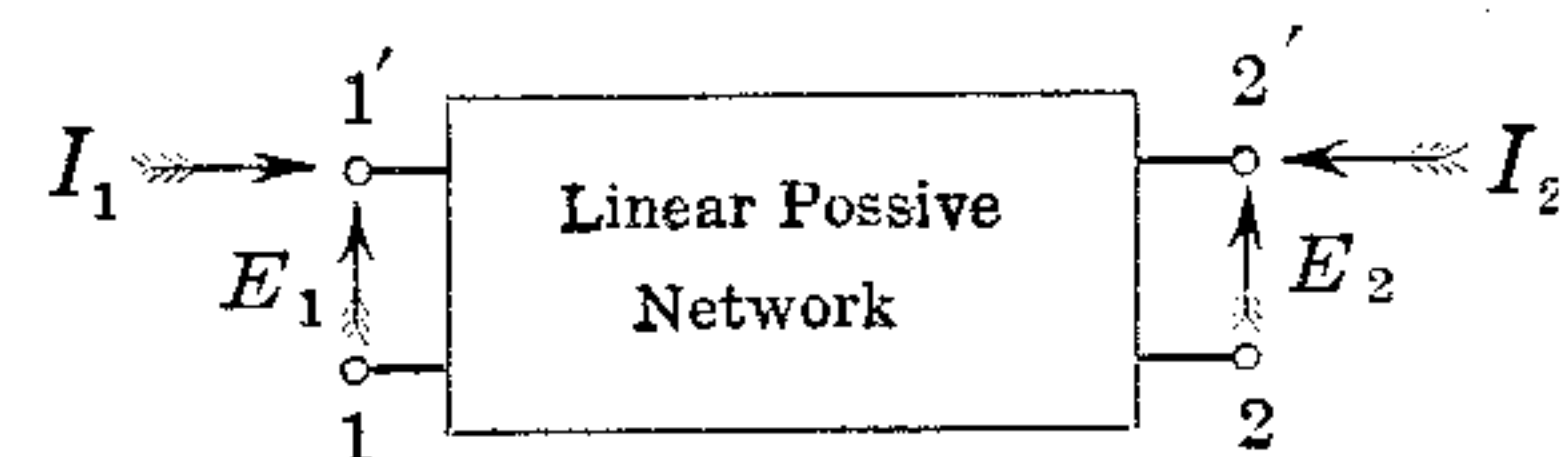


FIG. 32.—Schematic of a four-terminal network showing reference directions for voltage and current.

linear network having been assumed, it follows that these four quantities must be linearly related. It is also clear from our knowledge of the theory of a lumped-constant network that these relationships may be set up for any two quantities in terms of the other two. There are six such relations

in all, or more specifically, three pairs of mutually inverse relations. Namely, we may express

$$(a) \quad \begin{cases} I_1 \text{ and } I_2 \text{ in terms of } E_1 \text{ and } E_2 \\ E_1 \text{ and } E_2 \text{ in terms of } I_1 \text{ and } I_2 \end{cases}$$

or

$$(b) \quad \begin{cases} I_1 \text{ and } E_2 \text{ in terms of } E_1 \text{ and } I_2 \\ E_1 \text{ and } I_2 \text{ in terms of } I_1 \text{ and } E_2 \end{cases}$$

or

$$(c) \quad \begin{cases} E_1 \text{ and } I_1 \text{ in terms of } E_2 \text{ and } I_2 \\ E_2 \text{ and } I_2 \text{ in terms of } E_1 \text{ and } I_1. \end{cases}$$

All these modes of expression have their particular usefulness in the analysis of four-terminal network performance. We proceed now to derive these and to show their interrelations.

3. Derivation of fundamental relations. If we think of the voltages E_1 and E_2 as forming the closing loops to those meshes of the network which would be left open at the terminals 1-1' and 2-2' respectively, and designate these as meshes 1 and 2, then according to the method outlined in Chapter IV of Volume I, and taking due account of the reference directions assumed here, we may write

$$\begin{cases} I_1 = \frac{B_{11}}{D} E_1 - \frac{B_{12}}{D} E_2 \\ I_2 = -\frac{B_{21}}{D} E_1 + \frac{B_{22}}{D} E_2 \end{cases} \quad (273)$$

where D is the network determinant and $B_{ik} = B_{ki}$ are its first minors. Here it is convenient to introduce the notation

$$\left. \begin{aligned} y_{11} &= \frac{B_{11}}{D} \\ y_{12} &= y_{21} = -\frac{B_{12}}{D} \\ y_{22} &= \frac{B_{22}}{D} \end{aligned} \right\} \quad (274)$$

and then write instead of (273)

$$\begin{cases} I_1 = y_{11}E_1 + y_{12}E_2 \\ I_2 = y_{21}E_1 + y_{22}E_2 \end{cases} \quad (275)$$

If we invert these equations we obtain relations of the form

$$\begin{cases} E_1 = z_{11}I_1 + z_{12}I_2 \\ E_2 = z_{21}I_1 + z_{22}I_2 \end{cases} \quad (276)$$

The z 's may be found in terms of the y 's by simply solving the system (275) and comparing coefficients with those in (276). Writing for the determinant of (275)

$$|y| = \begin{vmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{vmatrix} = y_{11}y_{22} - y_{12}^2, \quad (277)$$

we find

$$\left. \begin{aligned} z_{11} &= \frac{y_{22}}{|y|} \\ z_{12} &= z_{21} = -\frac{y_{12}}{|y|} \\ z_{22} &= \frac{y_{11}}{|y|} \end{aligned} \right\} \quad (278)$$

The y 's and z 's may be interpreted as admittances and impedances of the network. Namely, if we suppose that the terminals 2-2' are shorted, then obviously E_2 becomes zero, and the equations (275) give

$$\left. \begin{aligned} y_{11} &= \frac{I_1}{E_1} \\ y_{21} &= \frac{I_2}{E_1} \end{aligned} \right\} \quad (279)$$

while if we imagine the end 1 shorted, then $E_1 = 0$, and we have

$$\left. \begin{aligned} y_{12} &= \frac{I_1}{E_2} \\ y_{22} &= \frac{I_2}{E_2} \end{aligned} \right\} \quad (280)$$